

Alinhamento Interativo de Ontologias usando Anti-Padrões de Alinhamento: Um Primeiro Experimento

Alternative Title: Interactive Ontology Alignment using
Alignment Antipatterns: A First Experiment

Jomar da Silva
Departamento de Informática
Aplicada - UNIRIO
Av. Pasteur 458,
Rio de Janeiro, Brasil
jomar.silva@uniriotec.br

Fernanda Araujo Baião
Departamento de Informática
Aplicada - UNIRIO
Av. Pasteur 458,
Rio de Janeiro, Brasil
fernanda.baiao@uniriotec.br

Kate Revoredo
Departamento de Informática
Aplicada - UNIRIO
Av. Pasteur 458,
Rio de Janeiro, Brasil
katerevoredo@uniriotec.br

RESUMO

O progresso nas tecnologias da informação e de comunicação tornou disponível uma grande quantidade de repositórios de dados, mas com uma grande heterogeneidade semântica, o que dificulta a sua integração. Um processo que tem sido utilizado para resolver este problema é o alinhamento de ontologias, que tenta descobrir as correspondências existentes entre as entidades de duas ontologias distintas, as quais por sua vez estruturam os conceitos que definem os dados armazenados em cada repositório. Existem várias abordagens na literatura para o alinhamento de ontologias, dentre as quais destaca-se a abordagem interativa, que considera a participação de especialistas para melhorar a qualidade do alinhamento final. Na abordagem interativa, uma questão importante é a eficiência da participação do especialista. Apesar dos avanços nos resultados obtidos na literatura, há ainda erros recorrentes nos alinhamentos obtidos pelas propostas de alinhamento interativo de ontologias, o que pode ser comprovado por uma iniciativa de avaliação conduzida anualmente pela comunidade científica (OAEI). Neste trabalho, é proposta uma estratégia interativa para o alinhamento de ontologias, denominada ALIN, que busca a aumentar a eficiência da participação do especialista através da utilização de anti-padrões de alinhamento. A estratégia ALIN foi avaliada considerando 3 anti-padrões disponíveis na literatura, e os resultados comprovaram o potencial da abordagem.

Palavras-Chave

Alinhamento de Ontologias, Ontologia, Alinhamento interativo de Ontologias, Anti-padrão de alinhamento.

ABSTRACT

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SBSI 2015, May 26–29, 2015, Goiânia, Goiás, Brazil.
Copyright SBC 2015.

A large amount of data repositories became available due to the advances in information and communication technology. Those repositories, however, are highly semantically heterogeneous, which hinders their integration. Ontology alignment has been successfully applied to solve this problem, by discovering correspondences between two distinct ontologies which, in turn, conceptually define the data stored in each repository. Among the various ontology alignment approaches that exist in the literature, interactive ontology alignment includes the participation of experts to improve the quality of the final alignment. An important issue for interactive ontology alignment proposals is to maximize the efficiency of the expert participation. According to OAEI (an evaluation initiative conducted annually), despite the advances in this field there are still recurrent errors in the final alignments obtained by the interactive ontology alignment proposals. In this paper we propose ALIN, an interactive ontology alignment approach that increases the efficiency of the expert participation through the use of alignment anti-patterns. ALIN was evaluated considering three anti-patterns available in the literature, and the results proved the potential of the approach.

Categories and Subject Descriptors

H.2.1 Logical Design - Data models

General Terms

Algorithms, Measurement, Design.

Keywords

Interactive Ontology Matching, Ontology, Correspondence Anti-pattern.

1. INTRODUÇÃO

O alinhamento de ontologias é uma operação importante em uma série de aplicações na área de sistemas de informação. Em parte destas aplicações o alinhamento de ontologias é feito na fase de projeto do software, podendo, assim, ser feito de maneira manual ou semi-automática[4]. Como exemplo destas aplicações temos a integração de ontologias, a integração de esquemas (de bancos de dados) tanto para a integração de bancos como para projetos de data warehouses. Em uma outra parte das aplicações o alinhamento de ontologias deve ser feito em tempo de

execução[4], como, por exemplo, na integração de web services. A proposta deste trabalho é semi-automática e se aplica às aplicações que podem ser feitas na fase de projeto.

O processo de alinhamento de ontologias tem como um de seus propósitos reduzir a heterogeneidade semântica entre representações distintas de um mesmo domínio. A existência de tal heterogeneidade se deve ao natural instinto humano de ter perspectivas diferentes e assim modelar situações semelhantes utilizando conceitualizações diferentes.

O processo de alinhamento de ontologias busca descobrir correspondências entre entidades relacionadas em ontologias [4]. Vários trabalhos já foram propostos na literatura para suportar este processo, entre eles se destacam os que seguem a abordagem interativa, que consideram a participação de um especialista no domínio que foi modelado pelas ontologias[6]. Esta abordagem tem alcançado resultados superiores às abordagens não-interativas, mas ainda há espaço para se alcançar melhores resultados [6].

Uma questão importante no alinhamento interativo de ontologias é a eficiência na participação do especialista. A eficiência na participação do especialista é medida pela quantidade de correspondência corretas achadas em relação ao número de perguntas feitas ao especialista.

Um anti-padrão de alinhamento, também chamado padrão problemático de alinhamento, é, mais precisamente, uma combinação de correspondências que geram uma insatisfabilidade de um conceito [7]. Guedes [1] aplicou anti-padrões de alinhamento para melhorar a qualidade do resultado de um processo de alinhamento de forma não interativa, retirando do alinhamento correspondências que levavam a inconsistências, de acordo com os anti-padrões.

Neste trabalho são avaliados os benefícios de considerar anti-padrões em uma abordagem interativa para o alinhamento de ontologias, denominada ALIN, com vistas a aumentar a eficiência na participação do especialista. Um primeiro experimento considerando 3 anti-padrões é descrito e seus resultados positivos mostram o potencial da abordagem ALIN.

Este artigo está estruturado como a seguir: As seções 2 e 3 explicam os conceitos envolvidos, a seção 4 mostra os trabalhos relacionados, o quinto mostra a estratégia ALIN, a sexta seção descreve a avaliação da abordagem através de um experimento e uma sétima seção conclui com limitações da abordagem e trabalhos futuros.

2. ALINHAMENTO DE ONTOLOGIAS

Uma ontologia tipicamente fornece um vocabulário descrevendo um domínio de interesse e uma especificação do significado dos termos neste vocabulário [4]. Uma ontologia é formada de entidades que podem ser: conceitos, relacionamentos ou atributos. Um modelo conceitual feito com o diagrama de classes da UML é uma ontologia. Um alinhamento de ontologias é o conjunto de correspondências, onde cada correspondência é uma relação (de equivalência, de generalização ou de disjunção) entre duas entidades destas ontologias.

Na Figura 1 pode-se ver um exemplo de um alinhamento de ontologias, onde as setas unindo entidades das duas ontologias são correspondências, e o símbolo acima delas a semântica da relação. No exemplo há duas ontologias (O1 e O2) acerca do domínio de Conferências. Em cada ontologia, os nomes dos atributos de cada conceito são precedidos por setas tracejadas e um relacionamento de especialização-generalização é representado por uma seta, que parte do conceito mais específico (por exemplo, *Book*) para o conceito mais geral (por exemplo, *Product*).

O alinhamento de ontologias é um processo cujo resultado final é um conjunto de correspondências descobertas. A Figura 2 ilustra uma visão geral das interfaces deste processo: a partir de duas ontologias de entrada (*O* e *O'*) que se pretende alinhar , é gerado um alinhamento *A'*. Opcionalmente, pode haver um alinhamento inicial *A* (em geral gerado por outra ferramenta de alinhamento) para servir como ponto de partida em busca de *A'*. O comportamento do processo de alinhamento é sintonizado por parâmetros (como o *threshold*, que representa um valor mínimo que deve ser excedido no cálculo da similaridade entre um par de entidades para que este par possa definir uma correspondência), Adicionalmente, o alinhamento pode levar em conta recursos externos, como por exemplo ontologias de referência ou repositórios de padrões e anti-padrões de alinhamento.

Para avaliar a qualidade do alinhamento gerado, duas medidas são comumente utilizadas, precisão e cobertura. Tais medidas são calculadas comparando-se o alinhamento gerado (do qual se quer avaliar a qualidade) com o alinhamento de referência (o conjunto de correspondências que são reconhecidamente verdadeiras). A precisão mede o quão correto é o alinhamento gerado, e é calculada como a razão entre o número de correspondências do alinhamento gerado que estão presentes no alinhamento de referência e o número total de correspondências do alinhamento gerado. A cobertura é calculada como a razão entre o número de correspondências do alinhamento gerado que estão presentes no alinhamento de referência e o número total de correspondências do alinhamento de referência. Há uma terceira medida, chamada medida-F, que é a média harmônica entre a precisão e a cobertura.

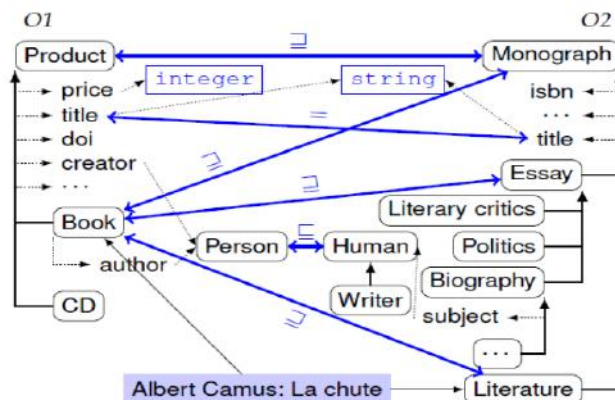


Figura 1. Exemplo de um alinhamento de ontologias [4]

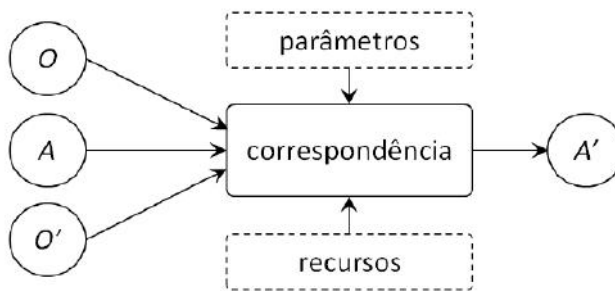


Figura 2. Processo de alinhamento de ontologias [4]

Durante o processo de alinhamento, a definição de uma correspondência entre duas entidades leva em conta a similaridade entre elas, que é calculada aplicando-se funções de similaridade. Existem diversas funções de similaridade na literatura, agrupadas segundo a perspectiva de análise que é considerada (terminológicas, sintáticas, semânticas, etc), como descrito por Shvaiko and Euzenat [4]. Em particular, no presente trabalho são utilizadas funções de similaridade terminológicas, tanto baseadas em strings (que consideram os termos que definem as entidades das ontologias como uma sequência de caracteres) como as linguísticas (que interpretam estes termos como objetos linguísticos).

Segundo Shvaiko and Euzenat [4], funções baseadas em strings comparam nomes e descrições das entidades das ontologias, considerando-os como sequências de letras em um alfabeto. Elas são baseadas tipicamente na seguinte intuição: quanto mais similares as strings, maior a probabilidade de tais entidades denotarem os mesmos conceitos na vida real. Usualmente, uma função de similaridade recebe um par de strings como entrada e retorna um número real entre 0 e 1, indicando a similaridade entre elas. Alguns exemplos de funções de similaridade baseadas em strings que são usadas em sistemas de alinhamento são: Jaro-Winkler [4], Jaccard [4] e n-Gram [4].

Já as funções de similaridade linguísticas utilizam recursos linguísticos (tais como dicionários ou *thesauri* específicos de domínio) para comparar palavras (nesse caso, o nome ou a descrição da entidade de uma ontologia é considerado uma palavra da linguagem natural). A similaridade entre duas palavras considera as relações linguísticas entre elas, por exemplo sinônimos, hipônimos e merônimos. Exemplos de funções de similaridade que utilizam recursos linguísticos são Jiang-Conrath [4], Lin [4] e Wu-Palmer [4].

Da mesma forma que com strings, usualmente as funções de similaridade linguísticas retornam um valor entre 0 e 1, sendo 1 quando ocorre a maior similaridade.

2.1 Anti-padrões de alinhamento de ontologias

Um anti-padrão de alinhamento é uma combinação de correspondências que geram uma insatisfabilidade de um conceito [7]. A Figura 3 ilustra alguns anti-padrões de alinhamento, onde as elipses são entidades de uma ontologia, a seta dupla em negrito representa uma correspondência e a seta mais clara identifica uma hierarquia, sendo o lado da ponta da seta a superclasse. O nome $oX:eX$ indica que a entidade eX pertence à ontologia oX .

Um anti-padrão de alinhamento reúne um conjunto de correspondências que, em conjunto com os relacionamentos das próprias ontologias (hierarquia de conceitos, por exemplo), necessariamente gera uma inconsistência lógica no alinhamento. Ou seja, dado um alinhamento de ontologias A que instancia um anti-padrão de alinhamento, a consistência lógica de A só pode ser garantida se um subconjunto não vazio das correspondências em A for removido. O alinhamento da Figura 3(c1), por exemplo, é inconsistente logicamente porque cada instância da classe $o1:e1$ é também instância da classe $o2:e1$ (por causa da equivalência entre estas classes), e cada instância de $o1:e1$ também é instância de $o1:e2$ (pelo fato de $o1:e2$ ser superclasse de $o1:e1$). Ainda, sabemos que cada instância de $o1:e2$ é também instância de $o2:e2$ (pela equivalência). Portanto, podemos deduzir que que as instâncias de $o2:e1$ são também instâncias de $o2:e2$, o que gera uma contradição lógica, pois $o2:e1$ e $o2:e2$ são disjuntas, ou seja, não há nenhum objeto do mundo real que simultaneamente instancie $o2:e1$ e $o2:e2$.

Algumas abordagens consideraram anti-padrões como recursos durante o processo de alinhamento de ontologias. Guedes definiu

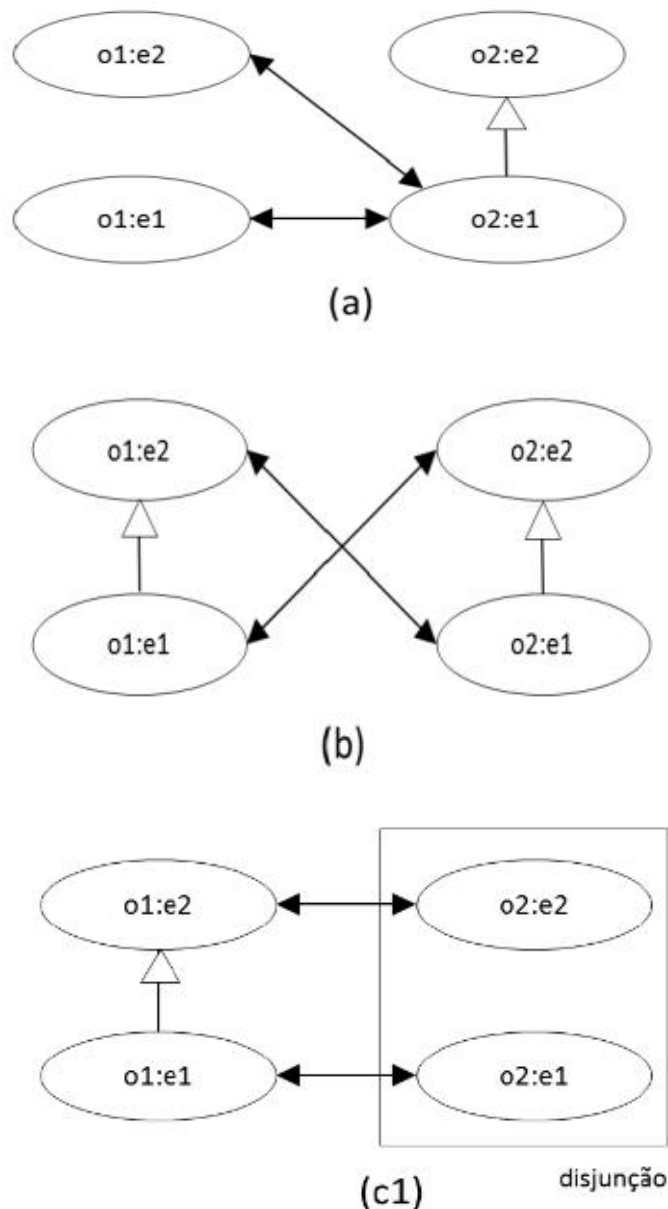


Figura 3. Anti-padrões de (a) correspondências de múltiplas entidades; (b) correspondências cruzadas e (c1) disjunção e generalização

empiricamente um conjunto de anti-padrões escavados a partir dos resultados de ferramentas avaliadas pela OAEI [8], e constituiu um repositório de anti-padrões de alinhamento. Adicionalmente, Guedes [1] propôs uma abordagem de alinhamento que submetia o alinhamento gerado para verificação, depois do processo de alinhamento já completamente executado. As correspondências deste alinhamento eram avaliadas para se verificar se instanciavam algum anti-padrão do repositório. Em caso afirmativo, uma das correspondências era retirada do alinhamento gerado e ele novamente submetido para verificação. O resultado do trabalho indicou que a utilização dos anti-padrões melhorou a qualidade do alinhamento. No entanto, os anti-padrões definidos são considerados por Guedes [1] apenas no

final do processo, e não de forma interativa, o que poderia antecipar a identificação de problemas em potencial durante a geração do alinhamento, aumentando a qualidade do alinhamento final e eliminando interações desnecessárias com o especialista.

3. CLASSIFICAÇÃO E CLUSTERIZAÇÃO

Classificação é a tarefa de associar objetos a uma de várias classes predefinidas [5]. A classificação é um processo de dois passos. No primeiro passo há um treinamento onde um algoritmo de aprendizagem constrói o classificador usando um conjunto de treinamento feito de tuplas e seus atributos de classe associados. Uma tupla é um vetor n-dimensional de atributos $X = (x_1, x_2, \dots, x_n)$. Cada tupla pertence a uma classe predefinida determinada por um outro atributo chamado atributo de classe, que tem valores discretos e não ordenados. Como o atributo de classe para cada tupla de treinamento é fornecido, este passo é conhecido como aprendizagem supervisionada.

Deve-se, então, estimar a acurácia de predição do classificador. Para isto se usa um conjunto de testes, com tuplas de teste com seus atributos de classe associados. Caso a acurácia encontrada seja satisfatória passamos para o segundo passo, ou seja, podemos utilizar o classificador para prever o atributo de classe de novas tuplas que venham surgir, e das quais não se conhece, a priori, seu atributo de classe. Como exemplos de classificadores temos o Perceptron [5], o Naive Bayes [5] e o Random Forest [5].

A definição do conjunto de treinamento é uma etapa crítica em um problema de classificação. Em alguns cenários, a base de treinamento é fornecida, ou segue-se uma abordagem de definição manual, onde o especialista de domínio classifica as instâncias que serão usadas para o treinamento. No entanto, na maioria dos casos reais isso não é factível, devido a restrições de disponibilidade dos especialistas para classificar um número de instâncias estatisticamente significativo. Uma das técnicas para selecionar automaticamente objetos para o conjunto de treinamento dentre um conjunto não classificado, é a query-by-committee [3]. Nesta técnica, a seleção dos objetos para a base de treinamento é feita de tal maneira que os classificadores acertem com mais acurácia a previsão das classes dos objetos não classificados. Os critérios para a escolha destes objetos se baseiam em um grupo de classificadores e em um conjunto bastante reduzido de treinamento inicial. A proposta é iterativa, pois quando um conjunto de objetos é classificado, este passa a fazer parte do conjunto de treinamento que o comitê utiliza para selecionar novos objetos para serem parte do seu novo conjunto de treinamento.

Já a análise de cluster divide objetos em grupos (clusters) que sejam úteis, significativos, ou ambos [5]. Diferente da técnica de classificação, que é supervisionada (ou seja, requer um conjunto de treinamento previamente classificado), a análise de clusters é não supervisionada. Ela agrupa objetos baseado somente nas características dos objetos e seus relacionamentos. A meta é que objetos em um grupo sejam similares (ou relacionados) uns aos outros e diferentes (ou não relacionados) a objetos nos outros grupos. Farthest-First é uma técnica de clusterização que divide o conjunto original em n (pré-definido) clusters, sendo retornado, nesta técnica, um objeto de cada cluster. É escolhido aleatoriamente um primeiro objeto, e a partir daí, até se alcançar n objetos, é escolhido outro objeto, que é o objeto mais distantes de todos os anteriores escolhidos. A Figura 4 ilustra os primeiros cinco passos da execução de um algoritmo Farthest-First. O primeiro ponto é escolhido arbitrariamente e cada ponto

sucessivo é o mais distante possível de todos os outros previamente escolhidos.

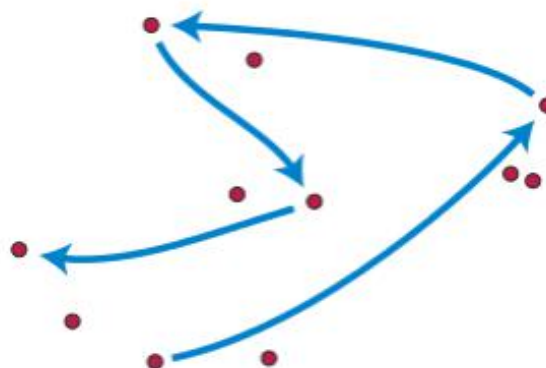


Figura 4. Ilustração da execução do algoritmo Farthest-First [wikipedia]

4. TRABALHOS RELACIONADOS

Lopes [2] propôs Jarvis, uma abordagem interativa para o processo de alinhamento de ontologias que aplica a técnica query-by-committee. Na abordagem Jarvis, a interação com o especialista é feita se mostrando uma correspondência a ele, e solicitando que ele responda se esta correspondência é correta ou não. O comitê definido na Jarvis é composto por 3 classificadores (Perceptron, Naive Bayes e Random Forest), e é utilizado em dois momentos no processo: na hora de se escolher quais correspondências vão ser apresentadas ao especialista para feedback e na hora de classificar automaticamente as correspondências que não foram classificadas pelo especialista. As correspondências a serem apresentadas ao especialista são aquelas que tem maior desacordo entre os classificadores, ou seja, aquelas em que dois classificadores dizem que pertencem ao alinhamento e uma diz que não ou vice-versa.

Jarvis considera cada par de entidades (possível correspondência) escolhida durante o processo como um objeto ou instância para os classificadores. Jarvis aplica técnicas de classificação no processo de alinhamento de ontologias, onde o atributo de classe indica se cada par de entidade pertence ou não ao alinhamento. Duas classes foram definidas: “sim” ou “verdadeiro” (para os objetos que são correspondências, e portanto farão parte do alinhamento gerado) e “não” ou “falso” (para os objetos que não são correspondências). Os demais atributos dos objetos são representados por valores das funções de similaridade (Jaro-Winkler, Jaccard, q-Gram, Jiang-Conrath, Lin e Wu-Palmer) para cada par de entidades (como ilustrado na Figura 5).

parDeEntidades
+Métrica Jaro-Winkler
+Métrica jaccard
+Métrica q-Gram
+Métrica Jiang-conrath
+Métrica Lin
+Métrica Wu-Palmer
+Pertence ao alinhamento

Figura 5. Par de entidades com seus atributos

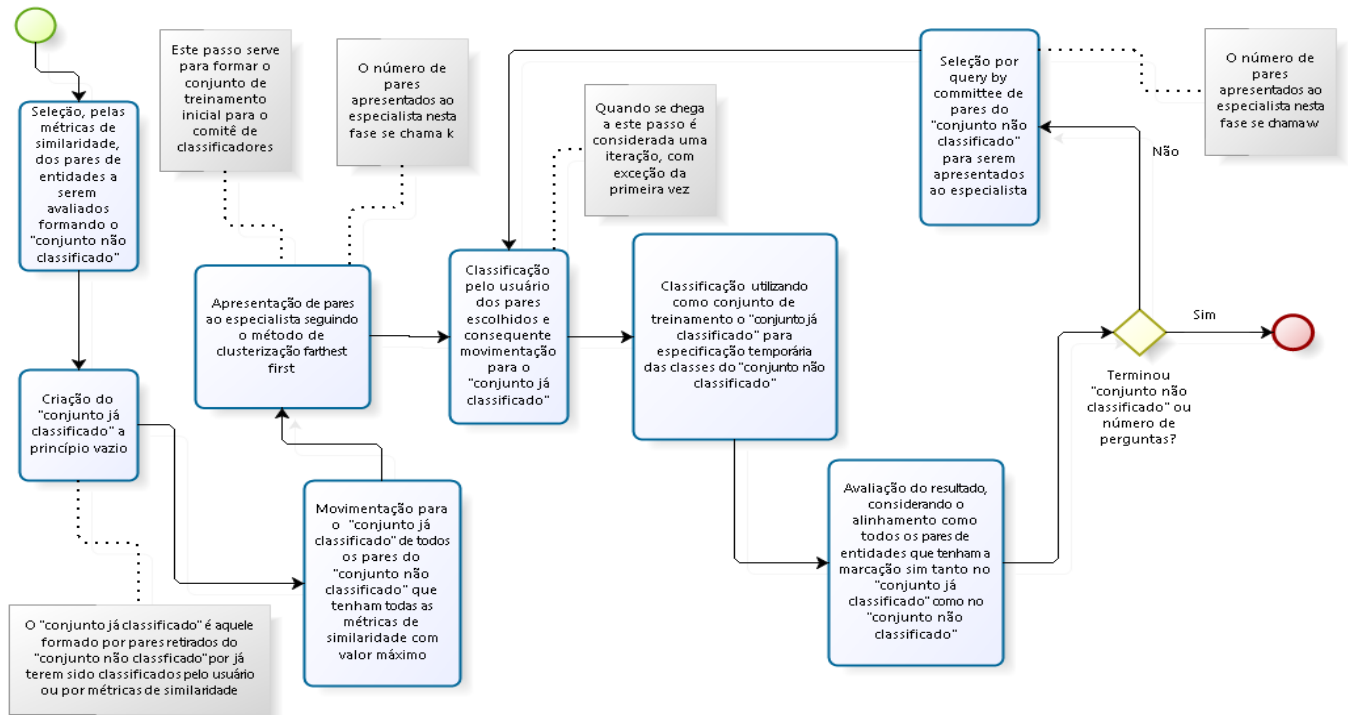


Figura 6. Processo interativo usando query-by-committee

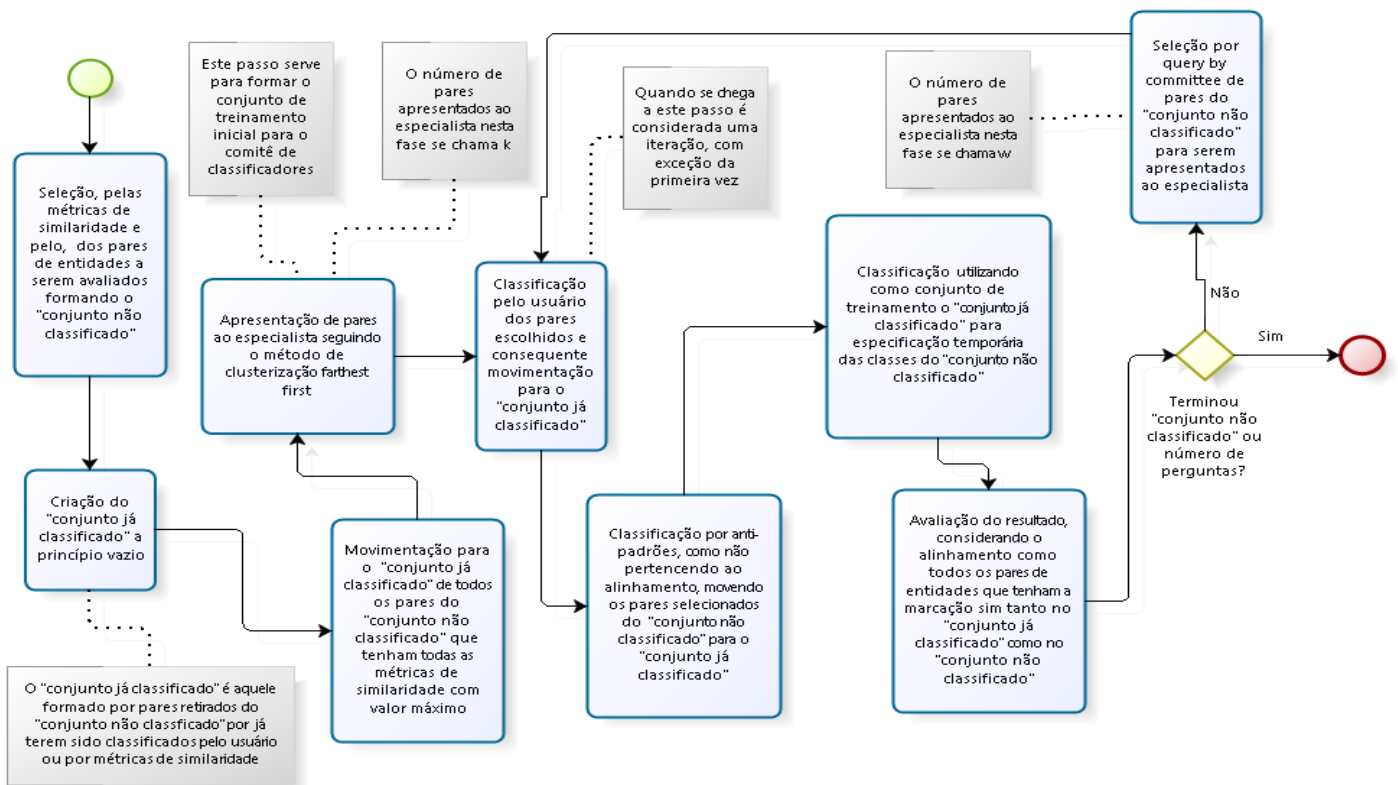


Figura 7. Processo interativo usando query-by-committee e anti-padrões de alinhamento

As instâncias que irão compor o conjunto de treinamento inicial são selecionadas aplicando-se a técnica Farthest-First apresentada anteriormente, sendo as correspondências os objetos retornados pelo clusterizador. A partir daí, Jarvis realiza várias iterações, onde em cada iteração escolhe e avalia correspondências candidatas. A abordagem Jarvis pode ser resumida da seguinte maneira: São aplicadas funções de similaridade terminológicas (Jaro-Wrangler, Jaccard, q-Gram, Wu Palmer, Lin e Jiang Conrath) para selecionar os pares com maior probabilidade de serem correspondências, dentre todos os pares possíveis de entidades entre as duas ontologias. Após esta escolha não são mais escolhidos pares dentro do total de possíveis, sendo avaliados pelo especialista somente estes pares escolhidos. Assim a cobertura máxima a ser alcançada no processo de alinhamento depende somente desta escolha inicial. A avaliação das correspondências escolhidas busca aumentar a precisão do alinhamento gerado, através da escolha de pares de entidades com similaridade máxima e da interação com o especialista, auxiliado pela técnica de query-by-committee, que ajuda o programa a escolher quais correspondências devem ser apresentadas ao especialista para classificação, e classifica correspondências ainda não classificadas pelo especialista. A abordagem Jarvis é mostrado ilustrada na Figura 6.

Outro trabalho relacionado é o de Guedes [1] que utiliza anti-padrões, mas de maneira não-interativa, utilizando os anti-padrões para melhorar um alinhamento já feito.

5. ALIN

Este trabalho propõe o Alin, uma estratégia interativa para o processo de alinhamento de ontologias que integra o uso de classificadores e clusterizadores como na abordagem Jarvis [2] com anti-padrões a cada interação com o especialista, estendendo portanto a proposta de Guedes [1].

A utilização dos anti-padrões se dá por ocasião da interação com o especialista. Caso o especialista responda “sim”, indicando que um par de entidades é uma correspondência, todos os pares incompatíveis com ele, de acordo com os anti-padrões, são classificados como “não”, indicando que estas correspondências não pertencem ao alinhamento. Isso amplifica as respostas dadas, pelo especialista, na interação, podendo-se saber a resposta para várias correspondências com somente uma pergunta ao especialista. A abordagem ALIN é ilustrada na Figura 7.

6. IMPLEMENTAÇÃO E RESULTADOS EXPERIMENTAIS

A abordagem proposta foi implementada utilizando as seguintes APIs em Java: Weka, com rotinas estatísticas e de KDD [10]; Simmetrics, com métricas de similaridade baseadas em strings [9]; WS4J, com métricas linguísticas baseadas na Wordnet [11]; e Alignment, que contém rotinas para manipulação de ontologias escritas em OWL [12]. A arquitetura da implementação do ALIN é ilustrada na Figura 8.

A avaliação da abordagem ALIN foi realizada considerando os resultados disponibilizados pela OAEI. A OAEI (Ontology Alignment Evaluation Initiative) é uma iniciativa internacional cuja meta é avaliar as ferramentas de alinhamento de ontologias existentes, publicando resultados comparativos entre elas. A OAEI disponibiliza alinhamentos de referência para diversos domínios de avaliação (*datasets*).

O *dataset* escolhido para a avaliação foi no domínio de conferências acadêmicas. O *conference dataset* é composto por 7 ontologias. Existe um alinhamento de referência entre cada par de

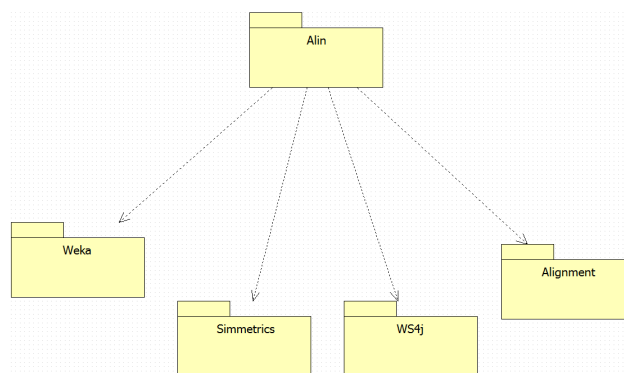


Figura 8. Arquitetura do Alin

ontologias, totalizando 21 alinhamentos de referência. Quanto à interação, cada feedback do especialista foi implementado através de uma consulta ao alinhamento de referência, que retorna se um par de entidades é uma correspondência ou não.

O experimento contemplou duas execuções, visando observar o impacto de se considerar anti-padrões no alinhamento interativo de ontologias. A primeira execução, sem a utilização dos anti-padrões, reproduziu a abordagem Jarvis [2]. A segunda execução seguiu a abordagem ALIN, incorporando os anti-padrões a cada interação com o especialista.

Em cada execução, as medidas de qualidade alcançadas representam a média dos valores encontrados nos alinhamentos gerados para cada par de ontologias do *conference dataset*.

6.1 Análise dos resultados

A abordagem ALIN foi avaliada segundo o impacto causado pela utilização dos anti-padrões na qualidade do alinhamento gerado e quanto ao número de interações necessárias com o especialista para a obtenção do alinhamento final.

O impacto da utilização dos anti-padrões na qualidade do alinhamento foi medido em termos de precisão, cobertura e medida-F. Foram feitas 2 execuções, sendo a primeira sem e a segunda com o uso dos anti-padrões. Em cada execução houve 5 iterações e foram definidos os parâmetros $k = 8$ e $w = 3$ (como definidos nas Figuras 6 e 7). Foram encontrados os resultados mostrados nas Figuras 9, 10 e 11.

O número de interações com o especialista também foi avaliado. Uma característica da utilização de anti-padrões é que, quando o especialista informa à abordagem ALIN que determinado par de entidades é uma correspondência, indiretamente informa que todos os outros pares inconsistentes com esta correspondência não pertencem ao alinhamento. Isto amplifica a propagação do conhecimento do especialista a cada feedback. Por exemplo, no alinhamento entre as ontologias *conference* e *ekaw* (pertencentes ao *conference dataset* da OAEI), foram selecionados 52 pares de entidades para serem avaliados pelo especialista. Na primeira execução, sem considerar anti-padrões, 36 pares são avaliados, sendo 23 diretamente pelo especialista e 13 por métricas de similaridade (Figura 12). Na segunda execução, quando são levados em consideração anti-padrões, dos 52 selecionados, 47 pares são avaliados, sendo 23 diretamente pelo especialista e 13 por métricas de similaridade (Figura 13), ou seja, 11 pares foram classificados observando-se os anti-padrões. Quando não se levou em consideração os anti-padrões, as medidas de qualidade foram: precisão = 0.722, cobertura = 0.52 e medida-F = 0.604. Já

utilizando anti-padrões, as medidas de qualidade foram: precisão = 0.736, cobertura = 0.56 e medida-F = 0.636.

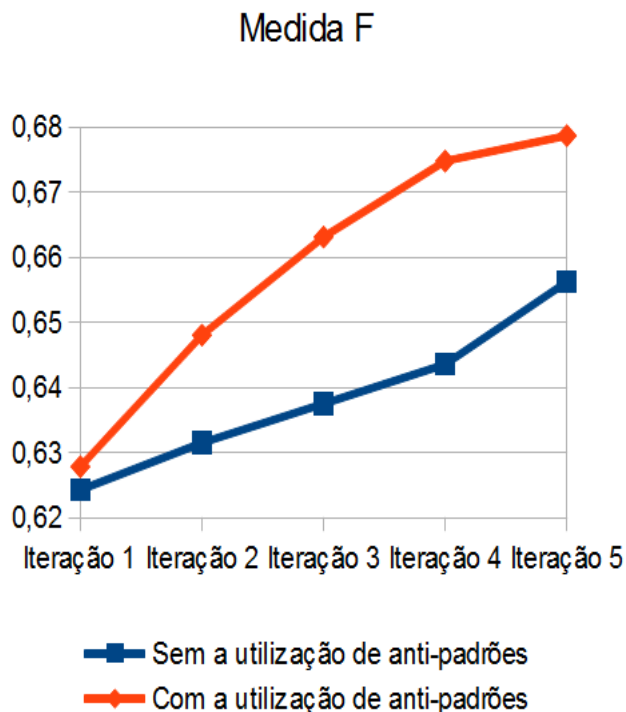


Figura 9. Valores de medida-F resultantes das execuções

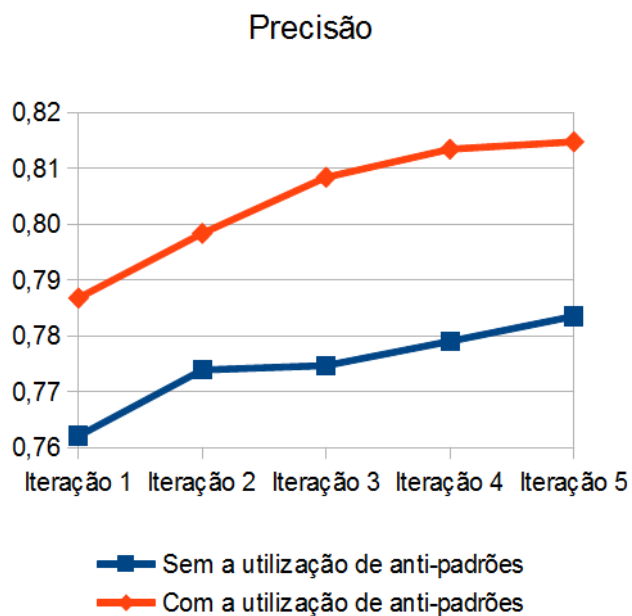


Figura 10. Valores de precisão resultantes das execuções

Cobertura

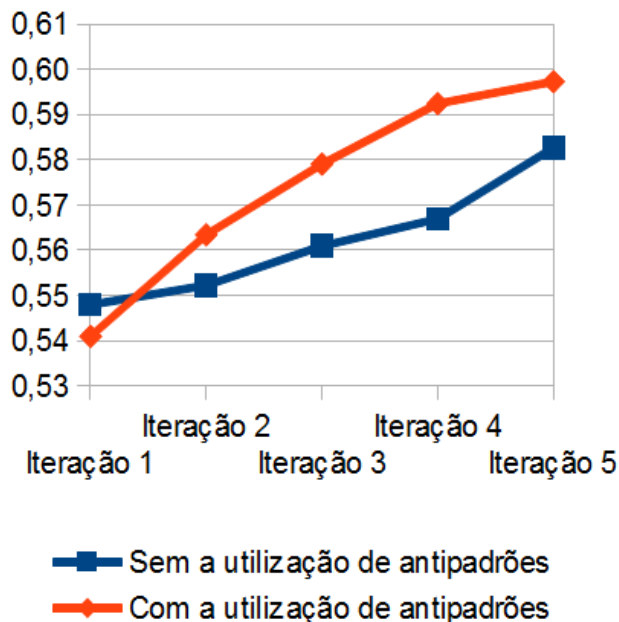


Figura 11. Valores de cobertura resultantes das execuções

	Número pares selecionados por similaridade	Número de perguntas respondidas pelo especialista	Número total de perguntas respondidas
Selecionadas por similaridade	13	0	13
Primeiro conjunto de perguntas	0	8	21
1ª iteração	0	3	24
2ª iteração	0	3	27
3ª iteração	0	3	30
4ª iteração	0	3	33
5ª iteração	0	3	36

Figura 12. Número de perguntas respondidas pelo especialista sem utilização de anti-padrões de alinhamento

	Número de pares selecionados por similaridade	Número de perguntas respondidas pelo especialista	Número de perguntas respondidas pela utilização de anti-padrões	Número total de perguntas respondidas
Selecionadas por similaridade	13	0	0	13
Primeiro conjunto de perguntas	0	8	10	31
1ª iteração	0	3	0	34
2ª iteração	0	3	1	38
3ª iteração	0	3	0	41
4ª iteração	0	3	0	44
5ª iteração	0	3	0	47

Figura 13. Número de perguntas respondidas pelo especialista com utilização de anti-padrões de alinhamento

7. CONCLUSÃO

Para avaliar se a utilização de anti-padrões aumenta a qualidade de um alinhamento de ontologias gerado por uma abordagem interativa, foi proposta a abordagem ALIN. A abordagem proposta foi avaliada sobre o *conference dataset* da OAEI, comparando os resultados obtidos contra os resultados da abordagem Jarvis, a qual não considera anti-padrões. Os resultados obtidos mostraram que a utilização de anti-padrões aumenta tanto a precisão como a cobertura do alinhamento obtido.

Uma outra observação importante é o aumento do número de pares de entidades avaliados pela abordagem, embora o número de interações com o especialista permaneça constante. A eliminação daqueles pares que não podem estar no alinhamento permite que o especialista responda somente às perguntas de correspondências que não são incompatíveis com outros pares já escolhidos. O aumento da qualidade também parece estar relacionado a isto. Alguns pares não selecionados anteriormente para avaliação podem ser agora avaliados, e alguns desses pares podem estar no alinhamento, embora isto possa não ser captado pelos classificadores.

A abordagem ALIN se diferencia dos trabalhos existentes na literatura em vários aspectos. Em Guedes[1] a utilização dos anti-padrões se dá somente ao final do processo de alinhamento, enquanto na abordagem ALIN proposta os anti-padrões são considerados a cada interação com o usuário. Ainda, com relação à proposta de Lopes[2], ALIN se diferenciou pela utilização de anti-padrões, eliminando das possíveis correspondências aquelas inconsistentes com as que foram positivadas pelo especialista. Os resultados do experimento mostraram que a sinergia entre o feedback do especialista e os anti-padrões contribuiu para

aumentar a qualidade do alinhamento e melhorar a eficiência da participação do especialista.

Como limitações da abordagem poderíamos citar que ela se baseia em ontologias representadas em OWL, que permite o registro de disjunções, que ocorre no anti-padrão mais utilizado na proposta, o de disjunção e generalização (c1 da Figura 3). Uma outra linguagem para representação de ontologias que não tivesse esta característica, como, por exemplo, um esquema de bancos de dados relacional, limitaria muito a abordagem.

Além disso a abordagem usa a hipótese de que o especialista nunca erra na resposta, o que nem sempre ocorre. Em caso de erro do especialista poderíamos ter uma propagação do seu erro com a utilização de anti-padrões.

Os anti-padrões de alinhamento podem ser utilizados de várias maneiras no alinhamento interativo de ontologias. Neste trabalho foram utilizados logo após cada interação com o especialista. Trabalhos futuros irão estudar a utilização de anti-padrões em outras fases do processo, como na escolha dos pares que serão apresentados ao especialista. Trabalhos futuros também irão considerar a inclusão de novos anti-padrões.

8. AGRADECIMENTOS

O projeto foi parcialmente financiado pela CAPES-DAAD (processo 4210-15-8) .

9. REFERÊNCIAS

- [1] Guedes, A., Baião F., Revoredo K. 2014. On the Identification and Representation of Ontology Correspondence Antipatterns - Department of Applied Informatics, Federal University of the State of Rio de Janeiro – UNIRIO, Brazil
- [2] Lopes V. 2014. Alinhamento Interativo de Ontologias: Uma Abordagem Baseada em Query-by-Committee – Msc Dissertation, Rio de Janeiro, RJ, Brazil: Universidade do Estado do Rio de Janeiro (UNIRIO), 2014
- [3] Prem M. and Raymond J. M. 2004. Diverse Ensembles for Active Learning. Proceedings of the 21st International Conference on Machine Learning, (ICML-2004), pp. 584-591, Banff, Canada, July, 2004
- [4] Shvaiko P. and Euzenat J. 2013. Ontology matching, Springer, Second Edition.
- [5] Tan, P., Steinbach M., Kumar V. 2006. Introduction to Data Mining – Addison Wesley
- [6] Paulheim H., Hertling S., and Ritze D. 2013. Towards Evaluating Interactive Ontology Matching Tools
- [7] Meilicke C., 2011. Alignment Incoherence in Ontology Matching
- [8] Guedes A., Baião F. and Revoredo K., 2014. Digging Ontology Correspondence Antipatterns
- [9] String Similarity Metrics for Information Integration, Sheffield University
- [10] WEKA Manual, 2015, The University of Waikato
- [11] <https://code.google.com/archive/p/ws4j/>
- [12] <http://alignapi.gforge.inria.fr/>