

MeMOTool: Uma Ferramenta para *Merging* Múltiplo de Ontologias

Fabiana F. de Araújo¹, Hélio R. Oliveira¹, Fernanda Lígia R. Lopes¹, Danusa R. B. da Cunha¹, Diana M. Oliveira², Bernadette F. Lóscio¹

¹ Departamento de Ciências da Computação – Universidade Federal do Ceará (UFC), Fortaleza, CE - Brasil

² Núcleo de Genômica e Bioinformática/NUGEN – Universidade Estadual do Ceará (UECE), Fortaleza, CE - Brasil

¹ {fabiana, heliorodrigues, fernanda.ligia, danusarbc, bernafarias}@lia.ufc.br,

² diana.magalhaes@uece.br

Abstract. *This paper introduces MeMOTool, a tool that aims to assist procedures of ontology merging. Its main core is the multiple ontology merging, which can be performed either randomly or through a clustering-based approach seeking to provide a high quality global ontology, the MeMO approach. MeMOTool is particularly useful on applications where an integrated view of several ontologies is required, while users will be able to manipulate (view, modify, merge, verify the quality of) a set of ontologies and their related alignments. MeMOTool is, then, one first step towards building a completely automatic ontology manipulation environment.*

Resumo. *Este trabalho apresenta a MeMOTool¹, ferramenta que objetiva dar suporte a operações de merging de ontologias. Sua principal funcionalidade é o merging múltiplo de ontologias que pode ser realizado de maneira aleatória ou seguindo a estratégia MeMO, uma estratégia baseada em clustering que busca gerar uma ontologia final de qualidade. A ferramenta tem o intuito de auxiliar as aplicações onde seja necessária uma visão integrada de várias ontologias. Usando MeMOTool, o usuário pode manipular (visualizar, alterar, combinar, analisar) um conjunto de ontologias e os alinhamentos relacionados às mesmas. A ferramenta visa contribuir no desenvolvimento futuro de um ambiente automatizado para manipulação de ontologias.*

1. Introdução

Uma ontologia caracteriza uma área ou um domínio de discurso, identificando conceitos e relacionamentos entre eles, geralmente definidos em uma linguagem formal, e disponibilizando metadados e informações explícitas quanto à semântica dos dados. No contexto da web semântica, o termo ontologia é usado como “uma especificação formal e explícita de uma conceitualização compartilhada” [Gruber, 1993]. As ontologias formam, portanto, a base da integração semântica e são consideradas fundamentais para

¹ <http://www.lia.ufc.br/~fabiana/memo.html>

o compartilhamento de informações. Porém, atualmente, há uma grande quantidade de ontologias disponíveis, sendo construídas de forma independente e, conseqüentemente, representadas de maneira heterogênea. Diante dessa diversidade de ontologias, há necessidade de combiná-las (*merging*) a fim de gerar uma ontologia que represente uma visão integrada de todas elas, visando facilitar a troca de informações. Assim, *merging* de ontologias é a criação de uma nova ontologia a partir de duas ou mais ontologias [Euzenat e Shvaiko, 2007].

Este trabalho apresenta a ferramenta MeMOTool que tem como foco principal o processo automático para *merging* múltiplo de ontologias [Araújo, 2009] utilizando a técnica de *custering* [Everitt, 1980] para se obter uma ontologia final confiável. As correspondências identificadas no início do processo, consideradas confiáveis por ter alguma interferência de um especialista, são reutilizadas para evitar o reprocessamento do *matching*. O processo de *matching* identifica os termos coincidentes das ontologias e, em seguida, são gerados seus mapeamentos. Este é um processo com custo elevado [Euzenat e Valtchev, 2003] que pode comprometer o desempenho da ferramenta. O intuito da ferramenta é facilitar o trabalho do usuário no que se refere às atividades de combinação.

Este artigo está organizado da seguinte maneira: a seção 2 apresenta outros trabalhos relacionados, no intuito de mostrar o contexto no qual a ferramenta MeMOTool está inserida; a seção 3 mostra a arquitetura e as principais funcionalidades da nossa ferramenta; a seção 4 explica o protótipo da ferramenta; a seção 5 destaca os trabalhos futuros e as considerações finais.

2. Trabalhos Relacionados

Muitas soluções já foram propostas para dar suporte às atividades relacionadas ao *merging* de ontologias, variando em soluções automáticas e semi-automáticas. O processo de *merging* pode ser realizado baseado em duas abordagens principais [Davies et al 2006]. A primeira, conhecida como *Merging* Completo, onde as ontologias dadas como entrada são combinadas e uma nova ontologia é gerada. Como exemplos de soluções que utilizam essa abordagem destacam-se: iPROMPT [Noy e Musen, 2002], COMA++ [Aumüller et al 2005] e ILIADS [Udrea et al., 2005]. Na segunda abordagem, as ontologias fonte são importadas e são definidos axiomas que especificam como os seus elementos estão relacionados. Como exemplos de soluções que utilizam essa segunda abordagem estão: Alignment API [Euzenat, 2004], HCONE-merge [Kotis et. Al. 2006] e OntoMerge [Dou et al., 2005].

Quando se pretende realizar o *merging* múltiplo de ontologias, contudo, a maioria das soluções propostas não atende a esse propósito, apesar de algumas possibilitarem, de uma forma aleatória, a extensão de suas soluções até a obtenção de uma ontologia global. Além disso, estas soluções dependem fortemente do processo de *matching* de ontologias, o que dificulta a realização do processo de combinação de forma automática, ou seja, sem a necessidade de interferência humana. Assim, objetivando estudar estratégias que possam realizar o *merging* múltiplo de ontologias, a ferramenta MeMOTool foi delineada e é apresentada na próxima seção.

3. A Ferramenta

A ferramenta MeMOTool foi desenvolvida inicialmente para validar a estratégia MeMO [Araujo et al 2009], uma estratégia para *merging* múltiplo de ontologias baseada em

técnica de *clustering* [Everitt, 1990], onde um conjunto de ontologias são combinadas de acordo com suas similaridades. A estratégia segue em duas etapas principais: (i) **construção da matriz de similaridade** onde será calculada a matriz diagonal de distância das ontologias, que representará o nível de similaridade entre elas. Para isso, serão usados os valores resultantes dos processos de *matching* entre todos os pares das n ontologias. Para a realização do *matching*, poderá ser usada uma das soluções já propostas na literatura (ex: CATO [Breitman ET AL 2005], COMA++ [Aumüller et al 2005]); (ii) **construção da árvore binária** onde será construída uma árvore binária que representará tanto as ontologias intermediárias, obtidas durante o processo de *merging*, como a ontologia global resultante. Para a geração dessa árvore será usada a matriz de similaridade gerada na etapa anterior e um algoritmo de *cluster* (agrupamento). Essa árvore será construída a partir dos nós folhas (ontologias locais) em direção ao nó raiz, onde estará a ontologia resultante do processo de *merging*.

Todavia, essa ferramenta visa não somente o uso dessa estratégia, mas também possibilitar que o usuário possa, em uma única ferramenta, executar o maior conjunto de atividades necessárias para manipular um conjunto de ontologias e alinhamentos voltados para o *merging* de ontologias. Desse modo, as principais funcionalidades disponíveis para o usuário nesta ferramenta são:

- *Verificar Qualidade*: para verificar a qualidade foi utilizada a ferramenta *OntEval*², que foi adicionada à *MeMOTool*. A ferramenta *OntEval* recebe como entrada duas ontologias e devolve como resultado um conjunto de valores que especificam as medidas de *Precision*, *Recall* e *F-Measure* (léxica e taxonômica), que caracterizam a “relação de proximidade” entre as ontologias analisadas [Dellschaft e Staab, 2006].
- *Merging de duas Ontologias*: duas ontologias iniciais são combinadas de acordo com o alinhamento entre elas, o qual é informado pelo usuário; uma ontologia resultante é retornada;
- *Merging Múltiplo (MeMO)*: várias ontologias são combinadas de acordo com a estratégia MeMO, onde um conjunto de alinhamentos referentes a essas ontologias são informados como entrada do processo. As combinações são realizadas duas a duas utilizando a função *Merging de duas Ontologias*, de acordo com a estratégia MeMO. O resultado final é uma árvore binária onde a raiz é a ontologia global e as folhas são as ontologias iniciais;
- *Merging Múltiplo (Aleatório)*: várias ontologias são combinadas aleatoriamente duas a duas, utilizando a função *Merging de duas Ontologias*, onde um conjunto de alinhamentos, que são referentes a essas ontologias, são informados como entrada do processo. O resultado final é uma árvore binária onde a raiz é a ontologia global e as folhas são as ontologias iniciais;
- *Junção de Alinhamentos*: um novo alinhamento é criado a partir de dois alinhamentos que compartilham uma ontologia comum [Araujo, 2009]. Essa função é utilizada para facilitar o processo automático de *merging* múltiplo, evitando repetição de chamadas ao processo de *matching*;

² <http://nlp.shef.ac.uk/abraxas/index.html>

- *Visualizar Ontologia*: uma ontologia pode ser visualizada através da interface da ferramenta, na qual suas classes e propriedades são exibidas;
- *Visualizar Alinhamento*: um alinhamento pode ser visualizado através da interface da ferramenta, permitindo a visualização de todo o seu conteúdo;
- *Gerar Alinhamento*: permite que o usuário crie os alinhamentos entre as ontologias diretamente pela MeMOTool (função em desenvolvimento).

A ferramenta MeMOTool foi desenvolvida em Java com destaque para a utilização de duas APIs: a Jena (<http://jena.sourceforge.net/index.html>) e a Alignment API [Euzenat, 2004] para manipulação dinâmica das ontologias (adição, remoção ou alteração de elementos) e geração e manipulação de alinhamentos respectivamente.

4. Experimentos

Para a validação da estratégia MeMO [Araújo, 2009] foram realizados experimentos considerando os seguintes tipos de ontologia: ontologias de referência (*gold standard*), ontologias resultantes do uso da estratégia MeMO e ontologias obtidas a partir de um processo de *merging* múltiplo aleatório. Durante os experimentos, as ontologias computadas a partir de processos de *merging* múltiplo (usando MeMO ou uma estratégia aleatória) foram comparadas com as ontologias de referência, a fim de que pudessem ser calculados os valores para as medidas de qualidade *Precision*, *Recall* e *F-Measure*. A principal medida levada em consideração foi a F-Measure, a medida harmônica da *Precision* e *Recall*. Pela Figura 1, pode-se notar que a estratégia MeMO foi melhor que a aleatória na maioria dos resultados. Em determinados casos, a estratégia aleatória gerou ontologias mais próximas da ideal, porém, foram resultados esporádicos e sem garantias. Ambas as estratégias tiveram seus resultados bastante comprometidos a partir da combinação de 16 ontologias, tendo a estratégia MeMO obtido resultados mais satisfatórios.

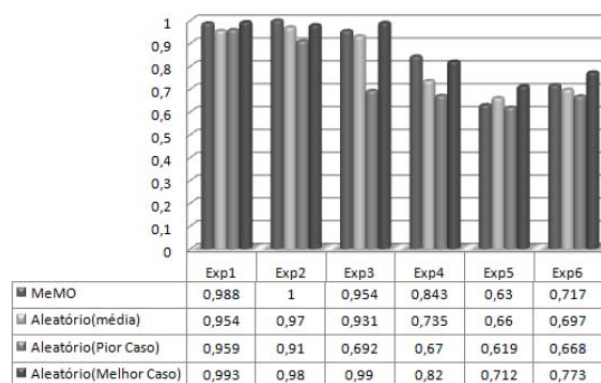


Figura 1. Resultados dos Experimentos

A Figura 2 ilustra a interface da ferramenta, onde em A é exibida uma lista de alinhamentos que podem ser adicionados, removidos ou visualizados pelo usuário. Em B é exibido o alinhamento (.rdf) selecionado em A, o qual também poderá ser editado posteriormente (essa tarefa de edição está em desenvolvimento). Em C, é exibida a árvore binária gerada durante o *merging* múltiplo de ontologias; os nós folhas representam as ontologias iniciais; os nós intermediários representam as ontologias geradas durante o

processo; o nó raiz representa a ontologia global. Cada ontologia da árvore binária pode ser visualizada em *D*.

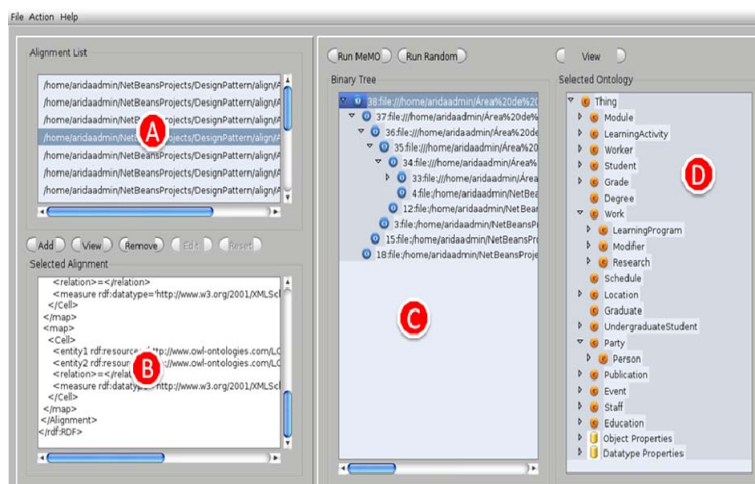


Figure 2. Interface gráfica da MeMOTool

5. Considerações Finais

A ferramenta demonstrada neste trabalho, a MeMOTool, foi desenvolvida com o objetivo principal de validação da estratégia MeMO [Araujo, 2009], uma nova solução para o *merging* de múltiplas ontologias com garantia de um resultado mais confiável que uma estratégia aleatória. O escopo da ferramenta foi ampliado para não se restringir apenas à validação da estratégia MeMO, mas também para abordar outras tarefas ligadas ao *merging* tais como o alinhamento entre ontologias e a combinação automática de duas ou mais ontologias. Essa ferramenta MeMOTool é o ponto de partida para uma ferramenta maior a ser posteriormente desenvolvida, a OnTools³, como um ambiente para alinhamento, integração e evolução de ontologias. Os próximos passos envolvem o aperfeiçoamento da ferramenta MeMOTool no sentido de melhorar a qualidade dos resultados, reforçar a usabilidade e agilizar a adição de novas funcionalidades.

Referências

- Araújo, F.F. (2009). “MeMO: Uma Estratégia para Merging Múltiplo de Ontologias baseada em *Cluster*. Trabalho de Mestrado. Universidade Federal do Ceará (UFC). Orientadora: Bernadette Farias Lóscio.
- Aumueller, D., Do, H.H., Massmann, S. and Rahm, E. (2005). “Schema and ontology matching with COMA++”. Proceedings 2005 ACM SIGMOD International Conference on Management of Data. Baltimore, MD, pages 906 – 908.
- Davies, J., Studer, R. and Warren, P. (2006). "Semantic Web Technologies: Trends and Research in Ontology-based Systems". John Wiley & Sons, 326 pp.

³ <http://arida.mdcc.ufc.br/wiki/index.php?n=Pesquisa.2008-OnToolsUmAmbienteParaAlinhamentoIntegracaoEEvolucaoDeOntologias>

- Dellschaft, K. and Staab, S. (2006). "On how to perform a gold standard based evaluation of ontology learning". In: Proceedings of ISWC-2006 International Semantic Web Conference. Athens, GA, USA. 2006.
- Dou, D., McDermott, D. and Qi, P. (2005). "Ontology translation on the semantic web". Journal on Data Semantics, 2: 35–57.
- Euzenat, J. and Valtchev, P. (2003). "An integrative proximity measure for ontology alignment". In: Proceedings of Semantic Integration Workshop. ISWC.
- Euzenat, J. (2004) "An API for Ontology Alignment". In: 3rd International Semantic Web Conference, ISWC 2004. Hiroshima (JP), Proceedings. Springer Verlag; 2004, pages 698-712.
- Euzenat, J. and Shvaiko, P. (2007). "Ontology Matching". Springer Verlag, 333 pp.
- Everitt, B. (1980). "Cluster analysis", 2nd ed., Gower Publishing Co.
- Gruber, T.R. (1993) "A translation approach to portable ontology specifications", Knowledge Acquisition, 6:199-221.
- Kotis, K., Vouros, G. and Stergiou, K. (2006). "Towards automatic merging of domain ontologies: The HCONE-merge approach". Journal of Web Semantics, 4: 60–79.
- Noy, N. F. and Musen, M. A (2002). "The PROMPT suite: Interactive tools for ontology merging and mapping", Technical report, SMI, Stanford University, CA, USA, 2002.
- Udrea, O., Miller, R.J. and Getoor, L. (2007). "Leveraging data and structure in ontology integration". In: Proceedings of SIGMOD 2007