

Abordagem para aquisição de conhecimento visual e refinamento de ontologias para domínios visuais*

Joel Luis Carbonera¹, Mara Abel¹, Claiton M. S. Scherer², Ariane K. Bernardes²

¹ Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

²Instituto de Geociências – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

{jllcarbonera,marabel}@inf.ufrgs.br, {claiton.scherer,ariane.kravczyk}@ufrgs.br

Abstract. *Em domínios visuais como a Medicina, a Meteorologia e a Geologia, as tarefas são realizadas através de uma aplicação intensiva do conhecimento visual dos especialistas. A natureza tácita do conhecimento visual impõe muitos desafios para Ciência da Computação em termos de aquisição, modelagem, representação e raciocínio. Neste trabalho será apresentada uma abordagem para aquisição de conhecimento visual e refinamento de ontologias em domínios visuais. Esta abordagem foi elaborada no contexto do projeto Obaitá, durante o processo de refinamento de uma ontologia de domínio para a Estratigrafia Sedimentar.*

1. Introdução

Domínios visuais são aqueles em que a resolução de problema inicia com um processo de casamento de padrões visuais, no qual são capturadas informações que serão utilizadas para suportar processos de inferência adicionais, em que são estabelecidas interpretações em um nível mais abstrato. Assim, domínios visuais são fortemente baseados na aplicação de conhecimento visual, que é o conjunto de modelos mentais que suportam o processo de raciocínio sobre informação relacionada ao arranjo espacial e outros aspectos visuais das entidades de domínio [Lorenzatti et al. 2009, Carbonera et al. 2011]. Estas características dos domínios visuais levantam muitos desafios para a Ciência da Computação, devido à natureza tácita do conhecimento visual.

Este trabalho insere-se no contexto do projeto Obaitá [Lorenzatti et al. 2009, Carbonera et al. 2011, Torres et al. 2011], desenvolvido pelo grupo BDI (grupo de bancos de dados inteligentes da UFRGS). Este projeto tem como objetivo a investigação de abordagens integradas para aquisição, modelagem, representação e raciocínio sobre conhecimento visual. Esta investigação tem sido conduzida no domínio da Estratigrafia Sedimentar. Espera-se que este projeto tenha como um dos seus resultados uma ontologia de domínio para Estratigrafia Sedimentar, que viabilize a construção de diversos sistemas baseados em conhecimento, que operem sobre uma mesma conceitualização do domínio.

Nesta fase do projeto, partimos de uma ontologia de domínio em desenvolvimento [Lorenzatti et al. 2009], e investigamos como esta ontologia deveria ser refinada para

*As bolsas de estudo deste projeto foram financiadas pelo programa PETROBRAS PFRH 17. Outros materiais foram fornecidos pela ENDEEPER Knowledge Systems)

atender às demandas levantadas pela tarefa de *interpretação dos processos deposicionais*. Com isso, nosso foco imediato é o refinamento da ontologia para atender às demandas de sistemas baseados em conhecimento que viabilizem a descrição visual dos objetos de domínio (em função da conceitualização utilizada pela comunidade)[Abel et al. 2012] e permitam a realização computacional da tarefa de *interpretação visual* de processos deposicionais geradores das fácies sedimentares inspecionadas [Carbonera et al. 2011]. É importante salientar que nesta abordagem buscamos realizar processamento de *descrições simbólicas* de feições visuais, de modo que o foco não é o processamento de dados visuais brutos (como imagens). Assim, assumimos que os usuários dos sistemas realizam o processo de abstração das feições visuais percebidas em representações simbólicas, ajustadas à ontologia de domínio.

A tarefa de interpretação de processos deposicionais depende do *conhecimento das feições visuais* de interesse e do *conhecimento inferencial* que possibilita que o especialista estabeleça a relação entre tais feições visuais (tomadas como evidências) e a interpretação correspondente. Uma vez que uma parcela considerável deste conhecimento visual é *conhecimento tácito* [Polanyi 1966], há uma resistência por parte dos especialistas em verbalizá-lo. Disto, segue-se que parte do conhecimento sequer é representado pela terminologia do domínio. Ou seja, parte dos fenômenos do domínio, presentes na conceitualização, podem não ter representações linguísticas. Com isso, as técnicas tradicionais para aquisição de conhecimento disponíveis na literatura (como *card sorting*, entrevistas estruturadas e não estruturadas, observação, limitação de informações, etc) não se mostram adequadas para realizar a aquisição de conhecimento visual [Abel 2001]. Este cenário motivou o desenvolvimento de uma abordagem de aquisição de conhecimento ajustada às necessidades deste projeto. Assumindo que a ontologia de domínio deve capturar o conhecimento necessário e suficiente para suportar a realização das tarefas de domínio, consideramos que a realização desta tarefa de interpretação pode oferecer uma boa ferramenta para avaliação da ontologia em desenvolvimento e, conseqüentemente, para revelar parcelas do conhecimento necessário faltante na ontologia em seu estado atual. Desta forma, utilizamos o próprio raciocínio especialista como uma ferramenta para atingir estes objetivos.

Na seção 2 será apresentado o domínio da Estratigrafia Sedimentar e seus principais objetos. Na seção 3 apresentaremos a nossa abordagem de aquisição de conhecimento visual e refinamento de ontologias. Na seção 4, apresentaremos um caso de uso desta abordagem o domínio. Na seção 5, apresentaremos algumas considerações finais.

2. Estratigrafia Sedimentar

A Estratigrafia Sedimentar é uma sub-área da Geologia que estuda as camadas que compõem a Terra e busca determinar a história da sua formação. Para alcançar este objetivo, o estudo inicia com a descrição visual de *corpos de rocha*. Estes, podem ser *testemunhos de sondagem*, que são cilindros de rocha retirados da subsuperfície terrestre por perfuração; ou *afloramentos*, que são exposições rochosas em superfície. A descrição destes corpos envolve discretizá-los em fácies sedimentares e descrever as características visuais (atributo visual e valor) relevantes de cada uma delas. A *fácies sedimentar* é uma dada porção de um corpo de rocha, visualmente distinguível das porções adjacentes. No domínio, assume-se que cada fácies observada é o resultado de um determinado *processo deposicional*, que é um evento que envolve a interação complexa entre forças naturais e

sedimentos. A partir da informação visual contida na descrição de cada fácies o especialista oferece uma interpretação de um possível processo deposicional responsável por gerá-la. Em passos subsequentes, estes processos deposicionais interpretados são utilizados para reconstruir a história geológica que resultou na formação geológica analisada.

3. Abordagem para aquisição de conhecimento visual e refinamento de ontologias para domínios visuais

A abordagem aqui proposta foi sugerida por duas constatações:

- Uma análise de casos (descrições visuais de fácies associadas à interpretação do processo deposicional gerador correspondente) disponíveis na literatura revelou informações visuais que não poderiam ser descritas com base na ontologia em foco, em seu atual estágio de desenvolvimento. Isto sugeriu que a ontologia não oferecia uma representação suficiente da conceitualização do domínio.
- Quando o especialista observa diretamente uma fácies *in loco*, ele tem à disposição todo o conhecimento visual que potencialmente pode ser obtido a respeito da fácies. Isto permite que ele realize a interpretação do processo deposicional formador da fácies, do modo mais adequado possível. Por outro lado, quando delegamos a tarefa de interpretação para um procedimento computacional, os únicos recursos disponíveis para processamento são as descrições simbólicas das feições visuais da fácies, oferecidas pelo geólogo, representadas computacionalmente de acordo com a ontologia de domínio.

A partir destas constatações, consideramos a hipótese de que submeter o especialista às mesmas limitações impostas à máquina quando ela realizaria esta mesma tarefa, permitira avaliar a ontologia, bem como revelar uma parcela importante da conceitualização do domínio que eventualmente não estaria contemplada por ela. Tendo isto em mente, concebemos uma *abordagem para aquisição de conhecimento visual e refinamento de ontologias de domínio guiada por resolução de problemas em contextos de informações limitadas*. Em termos gerais, esta abordagem utiliza o raciocínio de resolução de problemas realizado pelo especialista como um meio indireto para avaliar a suficiência da conceitualização representada na ontologia, oportunizando a identificação de lacunas de conceitualização e a posterior eliciação desta conceitualização junto ao especialista. Na Figura 1 a abordagem é representada de forma esquemática. Parte-se de um conjunto de *casos selecionados* da literatura (a) resultante de um processo de *seleção de casos*, previamente realizado. Para cada caso (b), realiza-se a *tradução* (3) da descrição visual do objeto (*não estruturada*) (b) para uma versão *estruturada* (c) (5), em função da *ontologia de domínio* (d). É importante salientar que este processo de tradução pode não preservar toda a informação da descrição original, uma vez que a ontologia de domínio pode não suportar parte da conceitualização necessária. A *descrição estruturada* (c) é submetida então ao *especialista* (f), para que ele realize a *interpretação* (4) desta informação. Neste ponto, interpretação refere-se ao processo de raciocínio utilizado pelo especialista para resolver o problema em foco. O produto desta etapa é uma *interpretação do especialista* (e), realizada a partir das informações limitadas em (c). A seguir, identifica-se (5) a *interpretação do caso* (g), a partir do caso descrito de forma não estruturada (b). Então, verifica-se a *correspondência* (6) entre a *interpretação do especialista* (e) e a *interpretação original* (g). No caso de haver correspondência, o processo termina. No caso de não haver correspondência, há uma chance da discordância ter ocorrido devido

a uma insuficiência da ontologia em representar uma parcela da conceitualização, inviabilizando a estruturação de informações relevantes contidas no registro original. Por esta razão, busca-se *identificar* (7) *conhecimento faltante* (h) na *ontologia de domínio* (d). Isto pode ser feito perguntando-se ao especialista que tipo de informação seria necessária para suportar a interpretação original; ou apresentando a descrição original ao especialista, de modo que ele possa identificar algum conceito necessário para ajustar informações que estão na descrição original, mas que não foram mantidos na descrição estruturada (devido a uma possível deficiência da ontologia). No caso de algum *conhecimento faltante* (h) ser identificado, ocorre o *refinamento* (8) da ontologia, através da incorporação deste novo conhecimento. O resultado do processo, ao fim da etapa (8), é uma ontologia de domínio refinada, em relação ao seu estágio inicial.

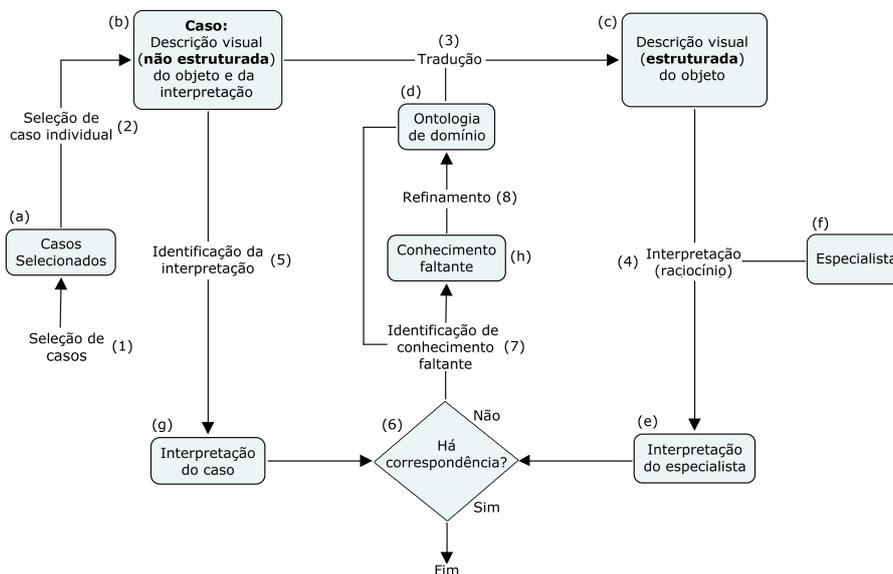


Figura 1. Representação esquemática da abordagem proposta

4. Aplicação da abordagem

Para realizar uma sessão de aplicação da abordagem com o especialista, primeiramente foram selecionados casos disponíveis na literatura, cada qual constituído por uma descrição das feições visuais de uma fácies sedimentar e de uma interpretação destas feições. Um exemplo de descrição selecionada da literatura é apresentado na Tabela 1, tal como ela está publicada (de forma não estruturada). Em um segundo passo, esta descrição foi traduzida para uma versão estruturada, usando os termos previstos pela ontologia em desenvolvimento, com o auxílio de dois geólogos em formação que conheciam a ontologia. Na Tabela 2 apresenta-se a versão estruturada. Neste estágio do desenvolvimento da ontologia, a versão traduzida contemplava apenas as informações que poderiam ser descritas com o atributo “Moda” e a “Estrutura Sedimentar”, as demais informações da descrição foram consideradas “Observações adicionais”, uma vez que não puderam ser descritas estruturadamente pela ontologia. Além destas informações, a Tabela 2 também apresenta a interpretação oferecida pela literatura, na coluna “Interpretação referencial”, enquanto na coluna “Interpretação do especialista” é apresentada a interpretação que o especialista realizou, sem ter acesso às informações descritas na coluna “Observações adicionais”. Na fase de verificação de correspondência, constatou-se uma diferença sutil entre

as interpretações. Na fase de verificação de conhecimento faltante, as informações rotuladas como “Observações adicionais” foram reveladas para o especialista, que constatou a relevância das informações não estruturadas para suportar a interpretação original (do caso selecionado). Dentre essas informações, a informação “Cascalho sustentado pelos clastos” revelou o atributo “Suporte de fábrica”, do qual “Suportado por clastos” é um dos seus valores; enquanto a informação “clastos comumente imbricados” revelou o atributo “Orientação de Fábrica”, do qual “Imbricado” é um dos valores. Uma vez revelados os atributos, o especialista listou outros valores possíveis para ambos. Com isso, realizamos o refinamento da ontologia, considerando estes dois novos atributos, e os seus respectivos valores possíveis. A Tabela 3 apresenta a mesma descrição de fácies, a partir da ontologia revisada. Comparando as Tabelas 1, 2 e 3, é possível notar que as informações tornam-se cada vez mais estruturadas ao longo das etapas, como consequência do refinamento da ontologia, através da incorporação de novos atributos e valores.

Descrição	Interpretação
Cascalhos sustentados pelos clastos, com estratificação horizontal pouco definida e clastos comumente imbricados	Barras longitudinais; depósitos residuais

Tabela 1. Exemplo de descrição de fácies encontrada na literatura. Descrição não estruturada

Moda	Estrutura Sedimentar	Observações adicionais	Interpretação referencial	Interpretação do especialista
Cascalho	Estratificação horizontal	Cascalho sustentado pelos clastos, clastos comumente imbricados, estratificação horizontal pouco definida	Barras longitudinais	Corrente trativa movendo cascalho, provavelmente barras longitudinais

Tabela 2. Exemplo de descrição de fácies submetida ao especialista. Estruturada de acordo com a ontologia em desenvolvimento

Moda	Estrutura Sedimentar	Suporte de Fábrica	Orientação de Fábrica
Cascalho	Estratificação horizontal	Suportado por clastos	Imbricada

Tabela 3. Exemplo de descrição de fácies estruturada a partir da ontologia revisada, com dois novos atributos para o conceito de fácies sedimentar

A utilização desta abordagem permitiu a identificação de 21 novos atributos com 62 novos valores. A comparação entre as versões da ontologia, antes e depois da aplicação desta técnica, fogem ao escopo deste trabalho, mas é apresentada em [Carbonera 2012];

5. Conclusão

Neste artigo apresentamos uma abordagem para aquisição de conhecimento visual e refinamento de ontologias para domínios visuais, bem como a aplicação desta abordagem no domínio da Estratigrafia Sedimentar. Esta abordagem apresentou bons resultados no que diz respeito à eliciação de novos atributos visuais descritivos. Este tipo de conhecimento é de fundamental importância em domínios visuais, visto que a tomada de decisão é fortemente dependente de caracterização visual qualificada dos objetos de domínio. Esta abordagem contribui para a aquisição de conhecimento e refinamento de ontologias porque, em geral, as abordagens disponíveis na literatura preocupam-se com identificação

de outros tipos de conhecimento (conceitos, relações, taxonomias, etc), sem focar-se na identificação de atributos. Assim, a utilização desta abordagem em conjunto com outras, pode promover a aquisição de um espectro mais amplo de tipos de conhecimento.

A abordagem ainda não foi adequadamente formalizada, uma vez que ainda precisa ser aplicada em outros domínios visuais, com o intuito de verificar sua generalidade. É possível que esta abordagem mostre-se adequada também em domínios não visuais. A verificação destes aspectos é um dos objetivos que pretendemos alcançar em trabalhos futuros. A abordagem apresentada pode ser considerada um resultado parcial do projeto Obaitá. É importante também salientar que a aplicação da abordagem pode revelar conceitos que não têm representação na terminologia do domínio, de modo que pode-se gerar termos que não são conhecidos pela grande comunidade da Geologia. No entanto, consideramos isto um aspecto positivo, no sentido de que pode-se revelar porções do conhecimento que ainda são difíceis de comunicar no domínio e sugerir termos que as representem. Estes termos revelados devem ser compreendidos como propostas, que devem ser avaliadas pela comunidade. O próximo passo natural é submeter a ontologia elaborada com a ajuda desta abordagem para a grande comunidade da geologia discuti-la em um processo de negociação de significados e construção colaborativa de ontologias [Torres et al. 2011].

Referências

- Abel, M. (2001). *Estudo da Perícia em Petrografia Sedimentar e sua Importância para a Engenharia do Conhecimento*. PhD thesis, Universidade do Rio Grande do Sul (UFRGS).
- Abel, M., Lorenzatti, A., Ros, L. F. D., da Silva, O. P., Bernardes, A., Goldberg, K., and Scherer, C. (2012). Lithologic logs in the tablet through ontology-based facies description. *AAPG Annual Convention and Exhibition*.
- Carbonera, J. L. (2012). Raciocínio sobre conhecimento visual: Um estudo em estratigrafia sedimentar. Master's thesis, Universidade Federal do Rio Grande do Sul (UFRGS).
- Carbonera, J. L., Abel, M., Scherer, C. M. S., and Bernardes, A. K. (2011). Reasoning over visual knowledge. In Vieira, R., Guizzardi, G., and Fiorini, S. R., editors, *Proceedings of Joint IV Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies*, volume 776.
- Lorenzatti, A., Abel, M., Nunes, B. R., and Scherer, C. M. S. (2009). Ontology for imagistic domains: Combining textual and pictorial primitives. In Heuser, C. A. and Pernul, G., editors, *ER Workshops*, volume 5833 of *Lecture Notes in Computer Science*, pages 169–178. Springer.
- Polanyi, M. (1966). *The tacit dimension*. Anchor Day Books, New York.
- Torres, G. M., Lorenzatti, A., Rey, V., da Rocha, R. P., and Abel, M. (2011). Collaborative construction of visual domain ontologies using metadata based on foundational ontologies. In Vieira, R., Guizzardi, G., and Fiorini, S. R., editors, *Proceedings of Joint IV Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies*, volume 776.