



20 e 21 de outubro
Instituto Nacional de Pesquisas Espaciais - INPE
São José dos Campos - SP

Uma Abordagem Centrada na Descrição de Dados para Melhoria do Processo de Composição Automatizada de Web Services Geográficos

Sérgio Ap. Braga da Cruz^{1,2}, Antonio Miguel Vieira Monteiro¹, Rafael Santos¹

¹Programa de Doutorado em Computação Aplicada – CAP
Instituto Nacional de Pesquisas Espaciais – INPE

²Embrapa Informática Empresa Brasileira de Pesquisa Agropecuária (Embrapa)
Caixa Postal 6041, CEP: 13083-886 - Campinas, SP - Brasil

{sergio@cnptia.embrapa.br, miguel@dpi.inpe.br, rafael.santos@lac.inpe.br}

Abstract. *This paper shows an approach for automated composition of geographic Web Services based in a deep geodata semantic description. Static and dynamic aspects of geodata exchanged among services in a composition are modeled by a geodata type ontology and geodata quality requirement rules, respectively. We use those models with a conditional planning strategy to produce more robust and reliable service compositions, capable to generate better quality geodata.*

Resumo. *Este artigo apresenta uma abordagem para automação do processo de composição de Web Services geográficos centrada na descrição semântica de geodados. Aspectos estáticos e dinâmicos dos geodados trocados durante as interações dos serviços de uma composição são modelados por uma ontologia de tipos de geodados e por regras de requisitos de qualidade de geodados, respectivamente. A construção automatizada de composições de serviços é realizada com base nestes modelos associados a uma estratégia de planejamento condicional. Esta abordagem permite a criação de composições de serviços mais robustas, confiáveis e capazes de produzir geodados de melhor qualidade.*

Palavras-chave: *composição automática de serviços, Web Services, qualidade de dados geográficos*

1. Introdução

O processo de composição de *Web Services* permite a obtenção de novos serviços com funcionalidades de alto-nível a partir de serviços componentes, implementados e distribuídos por diferentes instituições. Estes serviços compostos podem ser utilizados na automação de atividades de análise espacial e geoprocessamento, permitindo a utilização

integrada de geodados e procedimentos de análise que podem ser acessados por meio dos serviços componentes. A medida que o número de serviços componentes disponíveis se torna maior, mais difícil se torna o processo de construção manual de serviços compostos. As atividades de localização, seleção e integração dos serviços componentes se torna mais complexa.

Métodos de planejamento em IA são utilizados por algumas linhas de pesquisa que visam a automação do processo de composição [Rao and Su 2005]. Sycara et al [Sycara et al. 2003] apresentam as bases para utilização de descrições semânticas de serviços para seleção, invocação e gerenciamento da interação com serviços. Nestas soluções baseadas em planejamento, descrições semânticas dos serviços são elementos importantes para a automação e visam descrever a capacidade funcional de cada serviço. Descrições explícitas são realizadas por meio da associação do serviço componente a um conceito de um modelo semântico de descrições de operações de serviços. Descrições implícitas permitem a inferência da capacidade funcional de um serviço componente de maneira indireta, através da descrição semântica de seus parâmetros de entrada e saída, pré-condições de uso e efeitos de sua execução [Sycara et al. 2003, Sirin et al. 2004]. As propostas para composição de serviços apresentam diferentes variações para descrição implícita da capacidade funcional de serviços [Ye and Chen 2006, Constantinescu et al. 2005, Dong et al. 2007, Ko et al. 2008, Chafle et al. 2007]. Lemmens et al [Lemmens et al. 2006] e Yue et al [Yue et al. 2007] apresentam adaptações destas abordagens de composição para o contexto de geoprocessamento. Estas descrições com foco no serviço não modelam adequadamente aspectos da interação entre os serviços que ocorrem durante a execução de uma composição. Condições individuais dos geodados, encontradas durante o tempo de execução, podem inviabilizar ou comprometer os resultados produzidos por procedimentos de geoprocessamento e análise espacial. Estas condições podem ser inerentes ao processo de produção do geodado. Porém, a construção de novas composições de serviços define, dinamicamente, contextos não previstos de uso de um geodado, propiciando o surgimento destas incompatibilidades. A execução de composições de serviços construídas sem considerar a possibilidade de ocorrência destas incompatibilidades, pode gerar dados incorretos e de baixa qualidade quando comparados com os resultados obtidos pela execução interativa das atividades de geoprocessamento e análise espacial [Bailey and Gatrell 1995].

O presente trabalho apresenta uma abordagem para construção automatizada de composições de *Web Services* geográficos, capazes de tratar com estas incompatibilidades. Composições mais robustas, capazes de produzir geodados de melhor qualidade, são obtidas por meio do refinamento da descrição semântica de geodados trocados entre os serviços. Este refinamento na descrição dos tipos de geodados considera os seus aspectos estáticos e dinâmicos. O ponto de vista estático, corresponde a modelagem dos tipos de geodados utilizados em análise espacial e geoprocessamento para formação de uma ontologia. Neste trabalho esta modelagem organiza os tipos de geodados com relação às suas componentes espaciais e não-espaciais. A utilização de tipos parametrizados complementa o refinamento desta descrição permitindo a obtenção de uma ontologia concisa. O ponto de vista dinâmico, considera a definição de atributos de qualidade de geodados para caracterização das instâncias de geodados trocados entre os serviços durante a execução da composição. Estes atributos permitem a construção de regras para descrição

semântica de requisitos de qualidade de dados, as quais devem ser respeitadas para que ocorra a execução bem sucedida do serviço composto. Estes refinamentos nas descrições dos dados associadas a algoritmos de planejamento em IA para composição automatizada de serviços permite que sejam gerados serviços compostos mais confiáveis, capazes de identificar e contornar não-conformidades na qualidade dos dados trocados entre os serviços da composição.

2. Composição Automatizada de Web Services como Problema de Geração de Planos

Um problema de geração de plano consiste na descoberta de uma sequência de ações que permitem a mudança de um estado inicial do mundo para um estado final ou meta [Russell and Norvig 2003]. No contexto de composição de serviços, as ações correspondem aos serviços. Os tipos de dados dos parâmetros de entrada e saída dos serviços correspondem a estados do problema de geração de plano [Carman et al. 2003]. Estes parâmetros de entrada e saída definem a capacidade de transformação de estado de um *Web Service*, ou seja, implicitamente a sua funcionalidade [Sirin et al. 2004]. Neste trabalho assumimos que cada *Web Service*, realiza somente uma operação e que sua funcionalidade esta associada somente a capacidade de transformação de informação [Martin et al. 2007].

Desta forma, um problema de planejamento pode ser descrito como uma tupla $\langle D, D_{init}, D_{req}, W, \delta \rangle$, onde D é um conjunto de tipos de dados correspondente aos estados possíveis, D_{init} é um conjunto de tipos de dados disponíveis descrevendo os estados iniciais, $D_{req} \subseteq D$ é o conjunto de tipos de dados requisitados correspondente ao estado meta, W é o conjunto de serviços correspondendo às ações e $\delta : D \times W \rightarrow D$ é uma função de transição onde cada tupla representa implicitamente a semântica de um *Web Service*. A função

$$GEN(dout) = \{w \in W, din \in D \mid \exists \delta(din, w) = dout\}$$

é definida por conveniência, e fornece todos os *Web Services* que produzem uma instância de dados do tipo *dout*.

Um mecanismo para composição de *Web Services* encontra uma sequência de serviços que produzem instâncias de tipos de dados em D_{req} . Para isto o procedimento orientado por meta, apresentado na Figura 1, pode ser realizado. Este procedimento realiza uma busca reversa em largura para encontrar os serviços que produzem um determinado tipo de dado. Porém, serviços que apresentam um mesmo tipo de entrada e saída de dados não são incluídos adequadamente na composição. A capacidade funcional destes serviços não pode ser inferida. Alguns procedimentos em análise espacial e geoprocessamento tais como, ajuste de escala, interpolação de dados, filtragens e, suavização de taxas, se encaixam neste caso. Estes procedimentos atuam sobre propriedades de uma instância de geodados ao invés de atuar na transformação de tipos de dados, e desta forma não são incluídos no plano de execução de serviços.

O algoritmo da Figura 1, adota uma abordagem de planejamento clássica [Russell and Norvig 2003], e assume que os dados gerados na saída de um serviço são sempre adequados para uso por um serviço consumidor dos dados. Esta condição não

```

function COMPOSE(composite_request) returns a composition or failure
  IO-COMPOSE(composite_request, [])

function IO-COMPOSE(composite_request, path) returns a composition, or failure
{
  if(composite_request.Dinit contains composite_request.Dreq) then return empty composition;
  services ← PREDECESSOR(composite_request.Dreq)
  input_datatypes_set ← []
  for each service in services
    for each input in service.inputs
      add input.datatype to input_datatypes_set
  if(input_datatypes_set is on path) then return failure
  output_datatypes_set ← []
  for each service in services
    for each output in service.outputs
      add output.datatype to output_datatypes_set
  newcomposite_request.Dinit ← composite_request.Dinit
  newcomposite_request.Dreq ← (composite_request.Dreq ∪ input_datatypes_set
    - output_datatypes_set)
  plan ← IO-COMPOSE(newcomposite_request, [output_datatypes_set | path])
  if plan ≠ failure then return COMBINE(services, plan)
  return failure
}

function PREDECESSOR(requested_datatypes) return services
{
  services ← []
  for each datatype in requested_datatypes
  {
    candidate_services ← GEN(datatype)
    selected_service ← SELECT(candidate_services)
    services ← services + selected_service
  }
}

```

Figura 1. Procedimento para composição de serviços

é sempre encontrada em aplicações do mundo real. Dois tipos de falhas podem ocorrer durante a execução de uma composição envolvendo *Web Services* geográficos. Falhas operacionais, tais como serviços fora do ar ou inacessíveis, impedem o acesso aos geodados fornecidos pelo serviço. Um segundo tipo de falha é gerada quando um serviço de análise espacial recebe geodados inadequados para uso. Geodados com valores atípicos, distribuição espacial aleatória e imagens de sensoriamento com nuvens são exemplos de dados inadequados. Este trabalho trata deste segundo tipo de falha, relacionada com aspectos dinâmicos da interação entre serviços durante a execução de uma composição. Para isto, nós propomos a representação destas restrições na interação entre serviços como regras de requisitos de qualidade de geodados. Estas regras são utilizadas em conjunto com uma estratégia de planejamento condicional para automatizar a construção de composições de *Web Services* geográficos.

3. Descrição Semântica de Tipos de Geodados

A construção de uma composição de serviços baseada no casamento de tipo de dados de entrada e saída de serviços (Figura 1) é suportada por uma ontologia de tipos de dados. Esta ontologia descreve conceitos associados aos tipos de geodados e de seus inter-relacionamentos. A base dos algoritmos de encadeamento de serviços é a avaliação do relacionamento de subordinação de conceitos nesta ontologia [Paolucci et al. 2002, Guo et al. 2005, Klusch et al. 2006]. A descrição semântica dos parâmetros de entrada e saída dos serviços é realizada associando estes parâmetros a conceitos desta ontologia. Portanto, quanto mais precisa a descrição dos tipos de dados nesta ontologia, mais representativo será o conceito associado às entradas e saídas de um serviço. Com isto, melhor será o desempenho do algoritmo de encadeamento de serviços, resultando na seleção de

serviços mais adequados às tarefas requisitadas na composição.

Este trabalho utiliza uma ontologia de aplicação [Guarino 1998] para modelar conceitos de tipos de geodados do domínio de geoprocessamento e análise espacial necessários para automação da composição de serviços. A construção de ontologias é uma tarefa tratada na disciplina de Engenharia de Ontologias [Hitzler et al. 2009], e não possui uma solução única. Em nosso caso, assumimos que os tipos de geodados classificam instâncias de geodados formadas por um conjunto homogêneo de entidades apresentando um atributo espacial obrigatório e um atributo não-espacial opcional. Este conjunto define uma camada de informação [Tomlin 1990], unidade elementar de geodado trocada entre serviços. Além da descrição semântica de tipos dos parâmetros de entrada e saída de um serviço, a descrição semântica do papel destes parâmetros também é importante na identificação de possíveis encadeamentos entre serviços [Medjahed et al. 2003]. Em nossa abordagem os tipos de geodados são parametrizados para permitir a descrição da sua dependência semântica de outros tipos de geodado. Uma ontologia de atributos descreve quais tipos de atributos espaciais e não-espaciais podem ser associados a uma entidade. A Figura 2 ilustra como o tipo de dado espacial *SpatialRate* pode ser descrito em nossa ontologia.

Esta descrição é realizada por meio de restrições sobre propriedades gerais de tipos da ontologia. *SpatialRate* requer um parâmetro com papel de *SpatialDenominatorParameter* e um com papel *SpatialNumeratorParameter*. Estes dois parâmetros devem apresentar uma componente não-espacial do tipo *Numeric*. A componente não-espacial do tipo de geodado *SpatialRate* é representada pelo atributo *SpatialRate*, subclasse de *Numeric* (Figura 2(b)).

4. Aspectos Dinâmicos na Descrição Semântica de Serviços

O casamento de tipos de parâmetros de entrada e saída dos serviços não permite o encadeamento adequado de serviços que não promovem uma modificação de tipo dos dados. Para contornar esta restrição é necessário distinguir dois tipos de transformação de informação realizadas pelos serviços. A primeira transformação, tratada adequadamente com o uso de descrições semânticas estáticas dos serviços, corresponde a geração de um novo tipo de dado de saída a partir de um tipo de dado de entrada. O segundo tipo de transformação é realizada quando um serviço altera propriedades de um dado mantendo o seu tipo. Para permitir a inclusão de serviços que executam este segundo tipo de transformação é necessário descrever quando elas são necessárias e, como o procedimento da Figura 1 pode ser modificado para acomodar esta nova necessidade. Para isto, foram considerados os pré-requisitos para execução confiável de um serviço consumidor de geodados.

A realização confiável de alguns procedimentos em geoprocessamento e análise espacial é dependente de condições de uso do dado. Caso as condições de uso não sejam adequadas, em algumas situações é possível realizar procedimentos para sua adequação. Nós assumimos que estas condições definem requisitos de qualidade de dados, que podem indicar a necessidade ou não da execução de procedimentos de transformação de propriedades de um geodado. Uma vez que estas necessidades dependem de características individuais de cada instância de geodado, a decisão de execução ou não do serviço de transformação de propriedades somente poderá ser tomada durante a execução do serviço.

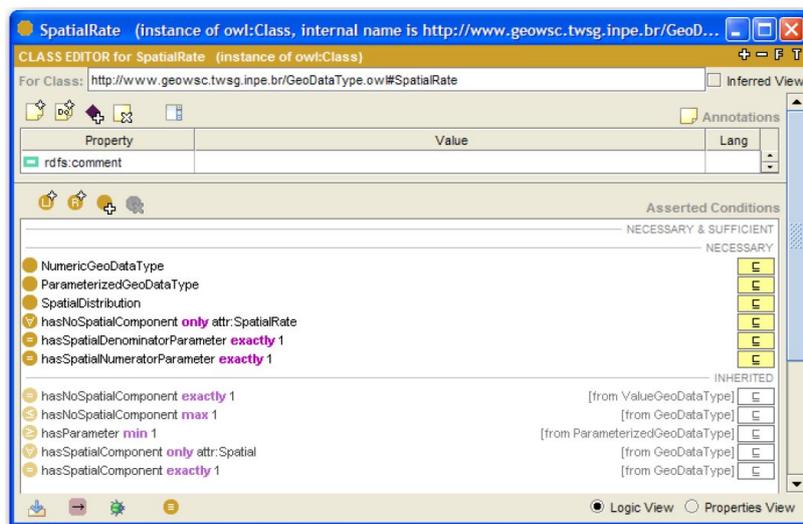
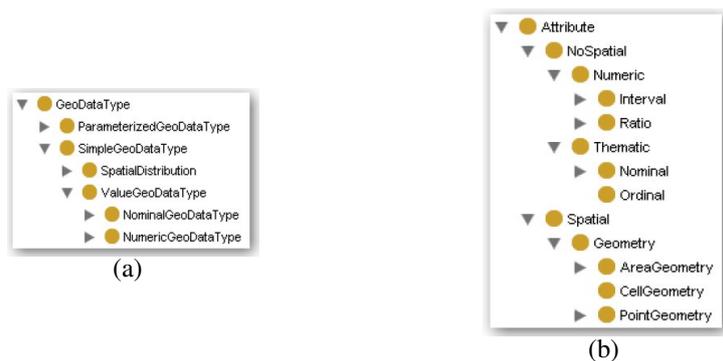


Figura 2. Descrição semântica de tipos de geodados. (a) ontologia de tipos de geodados; (b) ontologia de atributos; (c) tipo de geodado *SpatialRate*

Na composição, estes serviços são precedidos de testes de conformidade, necessários para avaliar as condições de uso dos dados. Estas condições são expressas em termos de atributos de qualidade de dados. Valores para estes atributos são calculados durante a execução da composição a partir de serviços que implementam procedimentos de análise espacial. Requisitos de qualidade dos geodados são descritos por regras semânticas expressas utilizando estes atributos, as quais determinam quais são os valores aceitáveis destes atributos que permitem o uso do geodado por um serviço. Nestas regras, atributos de qualidade que derivam seu valor diretamente da execução de um serviço são denominados indicadores de qualidade, e atributos que derivam este valor indiretamente, por meio de funções sobre outros atributos, são denominados parâmetros de qualidade. A Figura 3 ilustra regras em SWRL [Horrocks et al. 2004] descrevendo requisitos de qualidade de dados. As regras são expressas em termos de atributos de qualidade de dados organizados na taxonomia *DataQualityAttribute*.

5. Composição de Serviços baseados em Requisitos de Qualidade de Dados

A construção automatizada de uma composição de *Web Services* geográficos considerando os serviços de transformação de propriedades de geodados foi modelado como

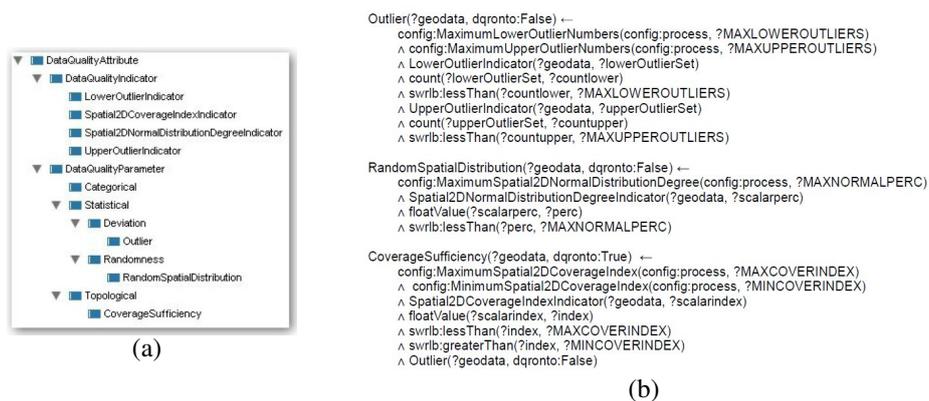


Figura 3. (a) Geodata Quality Attributes Ontology; (b) SWRL rules defining geodata quality requirements

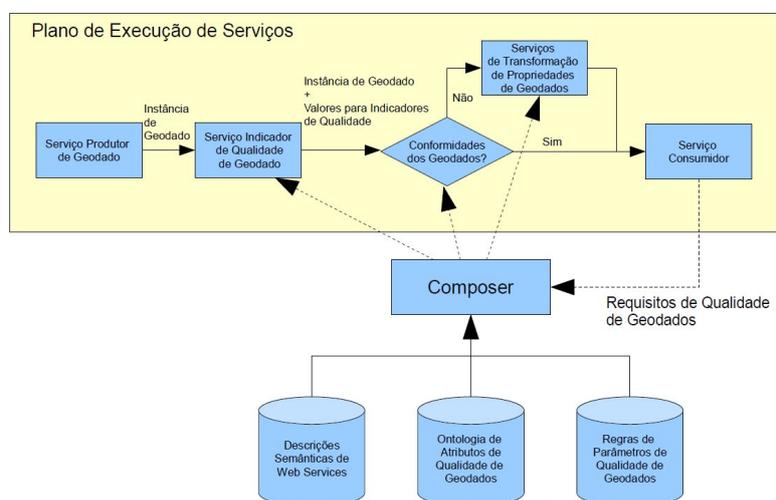


Figura 4. Participação de serviço indicador de qualidade e serviços de transformação de propriedades em uma composição

uma problema de planejamento condicional. O atendimento ou não aos requisitos de qualidade de geodados de um serviço consumidor define um conjunto de situações imprevisíveis no geodado gerado pelo serviço produtor, relativa ao atendimento a estes requisitos. Serviços de transformação de propriedades de geodados podem ser invocados baseados na ocorrência destas situações. Neste contexto, estes serviços funcionam como planos de contingência, adequando a instância do geodado às necessidades de um serviço consumidor. Os testes de conformidade funcionam como ações de sensoriamento, utilizadas na avaliação da existência de não-conformidades no geodado. A Figura 4 ilustra como serviços de transformação de propriedades e serviços indicadores de qualidade são incluídos em uma composição.

O mecanismo de composição prepara planos de contingência formados por serviços de transformação de propriedades para tratar possíveis ocorrências de não-conformidades. As expressões de teste de conformidade são construídas a partir das regras de parâmetros de qualidade e dos requisitos de qualidade, conforme mostrado na Figura 5. Através dos indicadores de qualidade presentes em uma regra é possível determinar quais

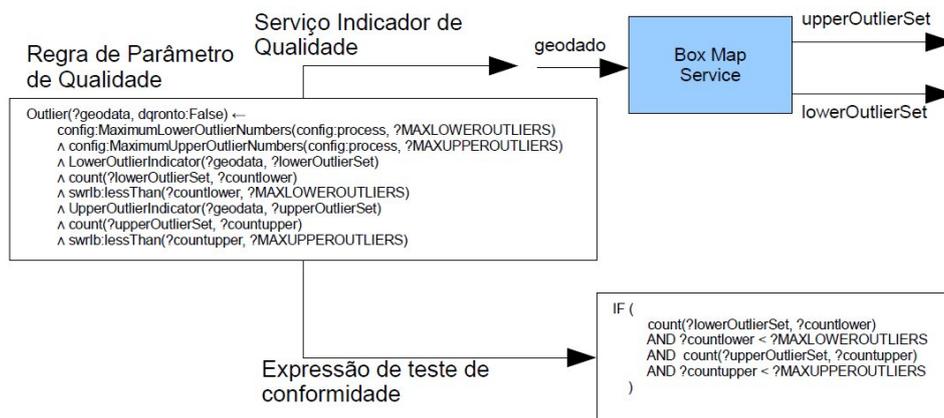


Figura 5. Exemplo de inferência de elementos da composição a partir de regra de parâmetro de qualidade de geodado

serviços indicador de qualidade deverão ser invocados.

Um protótipo do mecanismo de composição de serviços geográficos foi implementado utilizando a linguagem de programação Java e a ferramenta Jena [Jena 2010] para desenvolvimento de aplicações na Web semântica. A base de conhecimento do protótipo é formada por ontologia para tipos de geodados, ontologia de atributos de qualidade de geodados, regras de parâmetros de qualidade e descrição semântica de serviços. As ontologias e regras foram construídas respectivamente nas linguagens OWL [McGuinness and van Harmelen 2004] e SWRL por meio do editor de ontologias Protégé [Stanford Center for Biomedical Informatics Research 2010]. As descrições semânticas dos serviços são expressas na ontologia OWL-S [Martin et al. 2007]. O mecanismo de composição gera um processo simplificado em WS-BPEL, a partir da especificação de um tipo de geodado.

6. Resultados e Considerações Finais

A modelagem do problema de composição de serviços geográficos de acordo com a abordagem proposta permite a inclusão na composição, de serviços de transformação de propriedades e de serviços para extração de valores de indicadores de qualidade de maneira a tornar mais robusta o plano de execução de serviços gerado. Desta forma, o fluxo de execução de serviços da composição se aproxima da sequência de atividades de um especialista em análise espacial na resolução de um problema, executadas de maneira interativa com ajuda de SIG's e ferramentas de geoprocessamento [Bailey and Gatrell 1995]. Porém, a automação do processo de composição de serviços geográficos apresenta como vantagens a possibilidade de reuso de composições anteriormente geradas, de repetição da execução de procedimentos de análise e de sistematização destes procedimentos. O compartilhamento destes procedimentos e a organização de uma base de conhecimento de acordo como a proposta neste trabalho, permite o compartilhamento e padronização de critérios de análise entre especialistas do um domínio. Porém, para obter um uso efetivo da arquitetura SOA (Arquitetura Orientada a Serviços) na automação da execução de aplicações de geoprocessamento e análise espacial é necessário que a implementação dos *Web Services* geográficos, siga princípios de reuso, modularidade, autonomia, abstração e baixo acoplamento preconizados pelo desenvolvimento orientado a serviços [Erl 2004].

Esta proposta representa um passo inicial na utilização de parâmetros de qualidade na composição automatizada de serviços geográficos. A utilização de aspectos de qualidade de geodados relativos a resoluções espaço-temporais e escala de mapas no processo de composição precisam ser avaliados. A descrição das metas e o algoritmo para composição devem ser aperfeiçoadas para tratar restrições espaço-temporais.

O contexto de aplicação desta proposta assume a definição de uma base compartilhada de conhecimento dentro de um domínio específico de aplicação, visando a obtenção de resultados que possam ser utilizados em aplicações reais. A utilização de abordagens de mapeamento entre ontologias poderá ser realizada no futuro para integrar ontologias de diferentes domínios de aplicação.

Referências

- Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. London Longman Scientific and Technical.
- Carman, M., Serafini, L., and Traverso, P. (2003). Web service composition as planning. In *Proceedings of ICAPS'03 Workshop on Planning for Web Services*, Trento, Italy.
- Chafle, G., Das, G., Dasgupta, K., Kumar, A., Mittal, S., Mukherjea, S., and Srivastava, B. (2007). An integrated development environment for web service composition. *Web Services, IEEE International Conference on*, 0:839–847.
- Constantinescu, I., Binder, W., and Faltings, B. (2005). Flexible and efficient matchmaking and ranking in service directories. *Web Services, IEEE International Conference on*, 0:5–12.
- Dong, T., Li, Q., Zhang, K., and zhen Cui, L. (2007). An extended matching method for semantic web service in collaboration environment. In *CSCWD*, pages 508–513.
- Erl, T. (2004). *Service-Oriented Architecture : A Field Guide to Integrating XML and Web Services*. Prentice Hall PTR.
- Guarino, N. (1998). Formal ontology and information systems. In Guarino, N., editor, *Proceedings of the First International Conference on Formal Ontologies in Information Systems (FOIS-98), June 6-8, 1998, Trento, Italy*, pages 3–15. IOS Press, Amsterdam, The Netherlands.
- Guo, R., Le, J., and Xia, X. (2005). Capability matching of web services based on owl-s. In *DEXA Workshops*, pages 653–657.
- Hitzler, P., Krötzsch, M., and Rudolph, S. (2009). *Foundations of Semantic Web Technologies*. Chapman & Hall/CRC.
- Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., and Dean, M. (2004). Swrl: A semantic web rule language combining owl and ruleml. W3C Member Submission.
- Jena (2010). Jena a semantic web framework for java.
- Klusck, M., Fries, B., and Sycara, K. (2006). Automated semantic web service discovery with owls-mx. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 915–922, New York, NY, USA. ACM.

- Ko, J. M., Kim, C. O., and Kwon, I.-H. (2008). Quality-of-service oriented web service composition algorithm and planning architecture. *J. Syst. Softw.*, 81(11):2079–2090.
- Lemmens, R., Wytzisk, A., de By, R., Granell, C., Gould, M., and van Oosterom, P. (2006). Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing*, 10:42–52.
- Martin, D., Burstein, M., Mcdermott, D., Mcilraith, S., Paolucci, M., Sycara, K., McGuinness, D. L., Sirin, E., and Srinivasan, N. (2007). Bringing semantics to web services with owl-s. *World Wide Web*, 10(3):243–277.
- McGuinness, D. L. and van Harmelen, F. (2004). Owl web ontology language overview.
- Medjahed, B., Bouguettaya, A., and Elmagarmid, A. K. (2003). Composing web services on the semantic web. *The VLDB Journal*, 12(4):333–351.
- Paolucci, M., Kawamura, T., Payne, T. R., and Sycara, K. (2002). Semantic matching of web services capabilities. In *ISWC 2002: Proceedings of the First International Semantic Web Conference, Sardinia, Italy*, pages 333–347.
- Rao, J. and Su, X. (2005). A survey of automated web service composition methods. pages 43–54.
- Russell, S. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 2nd edition edition.
- Sirin, E., Parsia, B., and Hendler, J. A. (2004). Filtering and selecting semantic web services with interactive composition techniques. *IEEE Intelligent Systems*, 19(4):42–49.
- Stanford Center for Biomedical Informatics Research (2010). The protégé ontology editor and knowledge acquisition system.
- Sycara, K., Paolucci, M., Ankolekar, A., and Srinivasan, N. (2003). Automated discovery, interaction and composition of semantic web services. *Automated Discovery, Interaction and Composition of Semantic Web Services*, 1(1):27–46.
- Tomlin, C. D. (1990). *Geographic Information Systems and Cartographic Modeling*. Prentice Hall College Div.
- Ye, L. and Chen, J. (2006). Formal functional description of semantic web services: the logic description method. In *SOSE '06: Proceedings of the 2006 international workshop on Service-oriented software engineering*, pages 54–59, New York, NY, USA. ACM.
- Yue, P., Di, L., Yang, W., Yu, G., and Zhao, P. (2007). Semantics-based automatic composition of geospatial web service chains. *Comput. Geosci.*, 33(5):649–665.