

Composição de Web Services Geográficos baseada em Requisitos de Qualidade de Dados

Sergio A. B. da Cruz^{1,2}, Antonio Miguel Vieira Monteiro¹, Rafael Santos¹

¹ INPE Instituto Nacional de Pesquisas Espaciais
Av dos Astronautas, 1.758 Jd. Granja - CEP: 12227-010
São José dos Campos SP Brasil

² Embrapa Informática, Embrapa Caixa Postal 6041, CEP: 13083-886
Campinas, SP, Brasil

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

Abstract. *The geospatial Web Services composition allows implementing a geospatial process by integrating distributed geodata and procedures available in the institutions. AI planning methods can be used to automate these compositions. In this work we propose an improvement of this approach by using rule-based data quality requirements. These rules describe, at a semantic level, the quality requirements of messages exchanged among the composition Web Services. This approach allows us to produce Web Services composition, which can identify and overcome data quality nonconformities found in the data exchanged within the composition. This approach can create better reliability geospatial processes capable to produce better quality geodata.*

Resumo. *A composição de Web Services geográficos permite a implementação de processos geoespaciais por meio da integração de dados e procedimentos distribuídos por diferentes instituições. Métodos de planejamento em IA são uma das abordagens utilizadas na criação automatizada destas composições. Neste trabalho propomos um aperfeiçoamento desta abordagem com a utilização de regras para descrição semântica de requisitos de qualidade de dados. O uso destas regras permite a geração de serviços compostos capazes de identificar e contornar ocorrências de não-conformidades na qualidade dos dados trocados entre os Web Services da composição. Esta abordagem permite a criação de Web Services compostos mais confiáveis e capazes de produzir geodados de melhor qualidade.*

1. Introdução

Web Services [Booth et al. 2004] são componentes de software acessíveis por protocolos da Web, e que possuem autonomia para realização das funções para o qual foram projetados. O uso destes componentes permite que instituições possam fornecer acesso padronizado à seus dados e procedimentos, e desta forma compartilhá-los com as suas instituições parceiras. Neste caso, tanto o acesso aos dados quanto aos procedimentos internos da instituição são transformados em serviços acessados pela Web. A combinação destes componentes são a base para construção de aplicações na arquitetura SOA (*Service Oriented Architecture*) [Erl 2004]. A arquitetura SOA pode ser utilizada na automação de aplicações de geoprocessamento, as quais são caracterizadas pelo compartilhamento de

grandes volumes de dados e pela execução coordenada de atividades de análise geoespacial. Cada instituição pode criar os seus serviços de maneira independente, os quais podem ser compartilhados por diferentes processos. Para gerar uma composição de Web Services é necessário identificar quais dos serviços disponíveis são os mais adequados para realização de um determinado processo. Além disto, é necessário que sejam criados mecanismos para construção de um plano de execução e interação entre os serviços, de modo que eles possam, combinados, realizar os objetivos do processo.

Métodos de planejamento utilizando técnicas de IA são adotados em muitos mecanismos de inferência para composição automatizada de serviços. Nesta abordagem, os mecanismos de inferência geram um plano de execução de serviços a partir da descrição de uma meta. Em geoprocessamento, a meta pode ser expressa na forma de tipos de dados e restrições associadas (espaciais, temporais ou outras) a serem gerados pelo processo. Tendo a definição de uma meta, a base complementar para construção do processo é a descrição semântica dos serviços. Porém o foco somente na descrição dos serviços não garante que o processo obtido será capaz de atender de maneira efetiva a meta desejada. Os resultados gerados na execução de serviços compostos podem apresentar erros, decorrentes da existência de não conformidades com requisitos de qualidade de dados, nas mensagens trocadas entre os diferentes serviços produtores e consumidores de dados da composição. Cada instância de dado trocada entre os serviços possui características particulares que habilitam-no como um dado adequado para uso por um determinado procedimento em geoprocessamento. As características adequadas de uso podem ser descritas como requisitos de qualidade de dados de entrada de um serviço. O atendimento a estes requisitos durante a execução de um serviço composto é imprevisível, e pode ser visto como um não-determinismo relacionado às diferentes instâncias de dados trocados entre os serviços componentes. Serviços compostos construídos sem considerar estes fatores podem gerar respostas contendo dados de baixa qualidade e inadequados para o uso.

Neste trabalho descrevemos como requisitos de qualidade de dados geoespaciais podem ser incorporados a mecanismos para automação de construção de composições, de modo a minimizar a perda de qualidade de dados gerados pelos serviços compostos.

2. Composição Automatizada de Web Services

Várias soluções têm sido propostas para a automação da composição de serviços [Rao and Su 2005, Milanovic and Malek 2004, Papazoglou and Heuvel 2007], incluindo abordagens baseadas em workflow [Aggarwal et al. 2004, Sirin et al. 2005] e formalismo de Rede de Petri [Zhovtobryukh 2007, Hamadi and Benatallah 2003]. Algumas soluções propostas na literatura modelam o problema de composição de serviços como um problema de planejamento em IA. O longo histórico de pesquisas e estudos nesta área da IA fornece uma base sólida e importante para construção de soluções para diferentes tipos de aplicação [Russell and Norvig 2003], e será adotada como foco deste trabalho. Nas soluções por planejamento, a composição de serviços pode ser vista como um procedimento orientado por meta, que gera um plano de execução dos serviços a partir dos requisitos da composição. A capacidade funcional de cada serviço determina se o mesmo participará do plano de execução, e em caso positivo, como será a sua participação.

Duas abordagens principais são adotadas para implementar mecanismos que realizam a composição de serviços utilizando métodos de planejamento em IA. A pri-

meira considera que a capacidade funcional de um serviço pode ser caracterizada pela sua capacidade de transformação de informação, ou seja, da transformação de seus dados de entrada em dados de saída. Como requisito da composição é fornecido o tipo de dado acompanhado de restrições (espaciais, temporais, etc.) que se deseja obter com a execução da composição. Opcionalmente pode ser definido neste requisito, o tipo de dado disponível como entrada de toda a composição. O procedimento de construção da composição neste caso, consiste no encadeamento de serviços de modo a resolver as dependências de dados de entrada. O resultado do procedimento é um plano de execução de serviços que gera o tipo de dado requisitado, quando executado. Os trabalhos de Sycara et al [Sycara et al. 2003], Sirin et al [Sirin et al. 2004] e de Yue et al [Yue et al. 2007] utilizam variações desta abordagem.

Uma segunda abordagem para composição utilizando método de planejamento considera as pré-condições e efeitos do serviço, além da compatibilidade entre os tipos de dado de entrada e saída [Ye and Chen 2006, Constantinescu et al. 2005, Dong et al. 2007]. Pré-condições e efeitos são expressas na forma de proposições lógicas. A descrição do estado do modelo de mundo que habilita a execução do serviço, corresponde a sua pré-condição, e a mudança de estado promovida por esta execução corresponde ao efeito. O princípio do procedimento de composição consiste em encadear os serviços adequados de forma que, ao final, gerem o estado desejado. Os requisitos para composição correspondem neste caso a uma descrição do estado desejado e, opcionalmente, os tipos de entrada e saída da composição. A equivalência entre o estado desejado e o estado gerado por um serviço é realizada por meio de inferência lógica.

Nas soluções observadas na literatura, estas duas abordagens são utilizadas, considerando um ambiente de execução estático e determinístico, o que corresponde a aplicação da estratégia de planejamento clássico [Russell and Norvig 2003]. As composições resultantes da utilização desta estratégia podem ser corretas para um ambiente ideal, porém não são robustas para aplicação práticas. Composições de serviços considerando um ambiente ideal podem ter a sua execução corrompida devido a ocorrência de situações não previstas e desta forma produzir resultados incorretos. Um tipo de evento imprevisto está relacionado com as propriedades específicas das instâncias de dados trocadas entre os serviços. Por exemplo, um serviço de análise espacial pode receber geodados contendo valores atípicos (*outliers*). Neste caso o serviço poderá gerar resultados incorretos, e a propagação destes resultados pode corromper a execução de todo o processo. Desta forma, o usuário final do processo poderá receber respostas inadequadas para tomada de decisão. Neste trabalho nós tratamos deste tipo de evento.

3. Composição de Serviços baseada em Requisitos de Qualidade de Dados

Um serviço pode participar de uma composição como um produtor de dado ou como um consumidor de dados. Associado ao serviço consumidor de dados, podem ser definidos um conjunto de requisitos de qualidade de dados que determinam em quais condições uma instância de dado pode ser utilizada pelo serviço. Valores de indicadores de qualidade [Wang et al. 1993] das instâncias de dados devem estar disponíveis para serem confrontados com os requisitos de qualidades de dados exigidos por um serviço. A Figura 1 ilustra um esquema da interação entre um serviço produtor e um consumidor de dados.

Os requisitos de qualidade devem ser expressos com base em valores de indica-

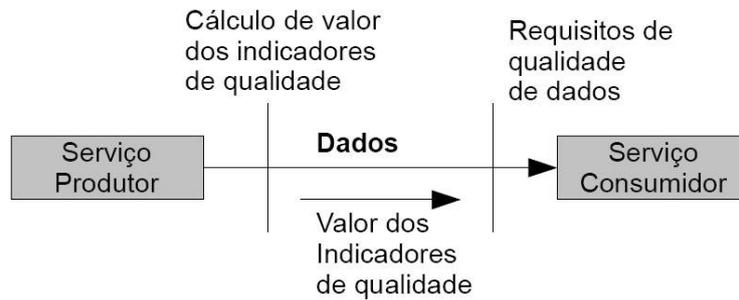


Figura 1. Esquema de interação entre serviços produtor e consumidor de dados

dores de qualidade que caracterizam dados adequados ao uso pelo serviço consumidor. Estes requisitos podem ser expressos na forma de regras combinando diferentes indicadores para formar um indicador agregado. As abordagens de Wang et al [Wang et al. 1993] e de Cockcroft [Cockcroft 1997] sugerem como podem ser definidos estes indicadores e regras.

Neste trabalho expressamos os requisitos de qualidade de dados semanticamente utilizando valores de indicadores de qualidade que caracterizam os dados adequados para uso por um serviço consumidor. Os indicadores fornecem uma métrica para estas características. Os atributos de qualidade de dados representam uma maneira para avaliar a conformidade com os requisitos de qualidade de dados fornecidos para o serviço consumidor. O valor destes atributos devem ser extraídos dinamicamente dos geodados, uma vez que eles não podem ser pré-calculados pelo serviço fornecedor. As não-conformidades identificadas indicam situações de uso inadequado de um serviço, e assim, devem promover a execução de algum plano de contingência.

Nós utilizamos o mecanismo de anotação semântica de dados geoespaciais baseado em regras [Klien 2007] para expressar os valores dos indicadores de qualidades de dados. As regras são construídas com base nestes indicadores. Serviços auxiliares dentro da composição são responsáveis pelo cálculo dos valores destes indicadores para os dados trocados entre os serviços produtores e consumidores. Estes serviços auxiliares são incluídos na composição durante a fase de planejamento. A partir da especificação dos requisitos de qualidade de dados definidos por um serviço consumidor é possível, conforme ilustrado pela Figura 2:

- Determinar quais indicadores de qualidade deverão ser utilizados e quais serviços serão necessários para extração dos valores destes indicadores para uma instância de dado geoespacial.
- Construir uma expressão para avaliação da aderência da instância de dado aos requisitos de qualidade.
- Determinar parâmetros para construção de um plano de contingência, a partir das propriedades da instância de dados que não atendem aos requisitos de qualidade.

A impossibilidade em prever se uma instância de geodado atenderá aos requisitos de qualidade pode ser modelada por meio de método de planejamento condicional. Neste caso, o serviço produtor produz dados atendendo ou não aos requisitos de qualidade de dados. Uma parte do plano de execução dos serviços trata as não-conformidades

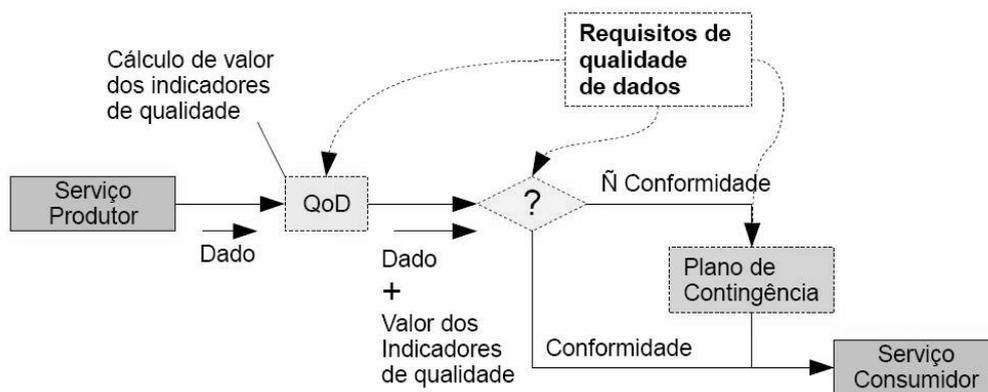


Figura 2. Usos dos requisitos de qualidade de dados no serviço composto

encontradas tornando, se possível, as propriedades não-conformes em conformes. A construção automática desta seção do plano é baseada nas propriedades não-conformes encontradas na instância de geodado.

4. Resultados e Considerações Finais

O protótipo de um mecanismo de inferência para composição de Web Services geográficos foi implementado utilizando uma base de conhecimento em Prolog, a qual armazena tanto os requisitos de qualidade de dados, quanto uma descrição semântica dos serviços. O mecanismo para composição foi implementado utilizando as linguagens de programação Java e Prolog, e pode produzir uma descrição simplificada de um processo na linguagem WS-BPEL [Alves et al. 2006]. O processo para geração de um Mapa de Moran [Druck et al. 2004] para ocorrências de Dengue por Municípios foi utilizado como estudo de caso. Para este caso o mecanismo para composição gera um processo a partir de requisitos da composição que descrevem, na forma de predicados Prolog, os tipos de dados a serem produzidos. As Figuras 3 e 4 mostram os diagramas dos processos gerados sem e com o uso dos requisitos de qualidade de dados, respectivamente. A Tabela 1 apresenta os serviços disponíveis para esta composição, e a Tabela 2 os requisitos de qualidade utilizados. O resultado obtido com a versão sem o uso dos requisitos de qualidade de dados esta sujeito às variações na qualidade dos dados. O Web Service que fornece dados sobre as ocorrências de Dengue pode fornecer geodados com distribuição espacial aleatória. Esta situação torna inviável o prosseguimento da análise espacial pelo processo. Valores atípicos (*outliers*) na taxa do número de ocorrências de Dengue por município corrompem os resultados gerados pelo serviço *MEMORAN*.

A versão do processo utilizando requisitos de qualidade de dados (Figura 4) contorna estes problemas incluindo na composição, serviços auxiliares e testes para detecção, durante a execução do processo, de geodados inadequados para uso. A detecção destes geodados direciona o fluxo de execução do processo para trechos de plano de contingência. Estes trechos visam transformar as propriedades dos geodados para permitir que eles possam ser utilizados no restante do processo. Se um plano de contingência não puder ser construído, o processo será capaz de informar sobre as ocorrências de não conformidades ao usuário final do processo. Os usuários, desta forma, poderão ter mais confiança nos resultados da execução do processo, evitando que geodados incorretos sejam propagados

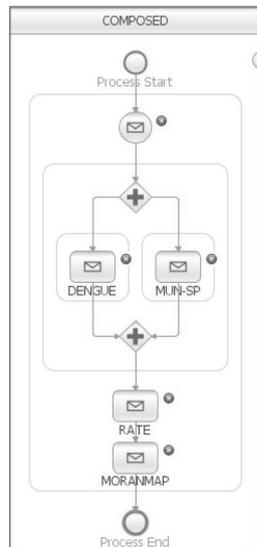


Figura 3. Processo frágil

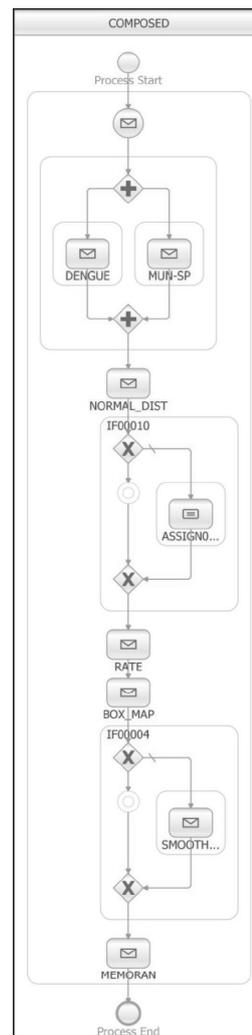


Figura 4. Processo robusto

para outras aplicações de geoprocessamento.

A composição de Web Services geográficos produzido utilizando esta abordagem trata adequadamente, em tempo de execução, os eventos inesperados relacionados com a existência de não conformidades com requisitos de qualidade de dados presentes nas mensagens trocadas entre os serviços da composição. Desta forma a execução do serviço composto será mais robusta, e possibilitará a geração de geodados de melhor qualidade. Como trabalho futuro, pretende-se complementar a base de conhecimento utilizada pelo mecanismo de composição para que sejam geradas descrições de processos em WS-BPEL operacionais. O mecanismo de composição deverá ser incorporado em um ambiente que permitirá o gerenciamento desta base de conhecimento. Nós esperamos que o conjunto de requisitos de qualidade de dados possa ser utilizado para formar uma base de conhecimento que possa ser compartilhada por diferentes composições de Web Services geográficos.

Tabela 1. Descrição dos Web Services usados no processo Mapa de Moran

Web Service	Descrição
DENGUE	Gera dados sobre eventos de Dengue
MUN-SP	Gera dados sobre divisão territorial do estado de São Paulo
RATE	Calcula a taxa de número de eventos por área
MEMORAN	Gera Mapa de Moran
SMOOTH_RATE	Reduz o número de valores atípicos de um conjunto de dados
BOX_MAP	Calcula número de valores atípicos
NORMAL_DIST	Mede o grau de distribuição espacial normal de um geodado

Tabela 2. Regras de Requisitos de Qualidade de Dados

Conceito	Regra SWRL
Não aleatoriedade	$norandomness(?data) \Leftarrow$ $normal_distribution(?data, ?perc)$ $\wedge lessThan(?perc, 5)$
Sem outliers	$nooutliers(?data) \Leftarrow upperoutliers(?data, ?upper)$ $\wedge lessThan(?upper, 5)$ $\wedge loweroutliers(?data, ?lower)$ $\wedge lessThan(?lower, 5)$

Referências

- Aggarwal, R., Verma, K., Miller, J., and Milnor, W. (2004). Constraint driven web service composition in meteor-s. In *SCC '04: Proceedings of the 2004 IEEE International Conference on Services Computing*, pages 23–30, Washington, DC, USA. IEEE Computer Society.
- Alves, A., Arkin, A., Askary, S., Bloch, B., Curbera, F., Golan, Y., Kartha, N., Sterling, König, D., Mehta, V., Thatte, S., van der Rijn, D., Yendluri, P., and Yiu, A. (2006). Web services business process execution language version 2.0. OASIS Committee Draft.
- Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., and Orchard, D. (2004). Web services architecture. Technical report, World Wide Web Consortium.
- Cockcroft, S. (1997). A taxonomy of spatial data integrity constraints. *Geoinformatica*, 1(4):327–343.
- 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 Shen, W., Yang, Y., Yong, J.,

- Hawryszkiewicz, I., Lin, Z., Barthès, J.-P. A., Maher, M. L., Hao, Q., and Tran, M. H., editors, *CSCWD*, pages 508–513. IEEE.
- Druck, S., Carvalho, M. S., Câmara, G., and Monteiro, A. M. V., editors (2004). *Análise Espacial de Dados Geográficos*. EMBRAPA.
- Erl, T. (2004). *Service-Oriented Architecture : A Field Guide to Integrating XML and Web Services*. Prentice Hall PTR.
- Hamadi, R. and Benatallah, B. (2003). A petri net-based model for web service composition. In *ADC '03: Proceedings of the 14th Australasian database conference*, pages 191–200, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.
- Klien, E. (2007). A rule-based strategy for the semantic annotation of geodata. *Transactions in GIS*, 11(3):437–452.
- Milanovic, N. and Malek, M. (2004). Current solutions for web service composition. *IEEE Internet Computing*, 8(6):51–59.
- Papazoglou, M. P. and Heuvel, W.-J. (2007). Service oriented architectures: approaches, technologies and research issues. *The VLDB Journal*, 16(3):389–415.
- 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. (2005). Template-based composition of semantic web services. In *In AAAI Fall Symposium on Agents and the Semantic Web*, pages 85–92.
- 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.
- 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.
- Wang, R. Y., Kon, H. B., and Madnick, S. E. (1993). Data quality requirements analysis and modeling. In *Proceedings of the Ninth International Conference on Data Engineering*, pages 670–677, Washington, DC, USA. IEEE Computer Society.
- 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.
- Zhovtobryukh, D. (2007). A petri net-based approach for automated goal-driven web service composition. *Simulation*, 83(1):33–63.