

APERFEIÇOAMENTO DO SISTEMA METEORO PARA GERENCIAMENTO DE DADOS PLUVIOMÉTRICOS NO RIO DE JANEIRO

Luan Soares de Andrade¹; Sergio Manuel Serra da Cruz²; Gustavo Bastos Lyra⁴ & Ednaldo Oliveira dos Santos^{4,5}

¹Bolsista FAPERJ, Discente do Curso de Sistemas de Informação/UFRRJ; ²Professor do DEMAT/ICE/UFRRJ, ⁴Professor do DCA/IF/UFRRJ; ⁵Professor orientador.

Palavras-Chave: Qualidade de dados, proveniência, precipitação.

Introdução

Uma das principais dificuldades enfrentadas por pesquisadores brasileiros quando se realiza trabalhos com dados meteorológicos é a obtenção de informações ordenadas, com as mesmas bases e em períodos similares, principalmente quando se pensa em trabalhar com dados climáticos, com séries de 30 ou mais anos de dados. Atualmente, percebe-se que no Brasil ainda não existe um sistema integrado em nível nacional disponível, com dados meteorológicos (VICENTE-SERRANO & BEGUERÍA-PORTUGUÉS, 2003; HEY *et al.*, 2009).

Além disso, inconsistências e variações nas séries climáticas existentes podem ocorrer devido a inúmeros motivos, por exemplo, erro humano, condições do instrumento de medição, processamento dos dados, variações no tempo de observações e às mudanças no ambiente circundante (VICENTE-SERRANO *et al.*, 2010).

Assim, não é suficiente apenas medir, é necessário também processar, corrigir, gerar e dar consistência aos dados coletados da maneira mais eficiente possível (WISSMANN, 2006). Isso se justifica porque a qualidade dos dados é um fator decisivo de sucesso. No entanto, somos muitas vezes confrontados com dados mal organizados, com falhas ou aparentemente irrelevantes. Para lidar com este tipo de problema, o primeiro passo deste processo foi o desenvolvimento da plataforma *Meteoro* como uma ferramenta de pré-processamento dos dados que assimila o formato dos dados produzidos por bases públicas disponíveis, além de organizar e tratar esses dados.

Entretanto, visto que o sistema *Meteoro* se tratou de uma proposta inicial, iniciou-se um aperfeiçoamento de suas características a fim de tornar o Sistema mais abrangente e robusto para serem usados posteriormente em vários estudos e aplicações, principalmente nas áreas de climatologia e ambiental.

Material e Métodos

Os dados utilizados nesta pesquisa fazem parte de séries de dados meteorológicos coletados em diversas estações pluviométricas distribuídas no estado do Rio de Janeiro (latitudes 20° 45' e 23° 21' S e longitudes 40° 57' e 44° 53' W). As séries climáticas de precipitação pluvial foram obtidas em arquivos de textos disponíveis no banco de dados da Agência Nacional de Águas (ANA), com auxílio da ferramenta HIDROWEB (<http://hidroweb.ana.gov.br>). Esses dados são obtidos de arquivos textos estruturados, e incorporados ao sistema de pré-processadores *Meteoro* (LEMOS FILHO *et al.*, 2013).

O sistema de pré-processadores *Meteoro* foi desenvolvido e parte integrante do projeto inicial que contemplou (i) um sistema de gerenciamento de banco de dados meteorológicos; (ii) rotinas computacionais para pré-processar os dados; (iii) preenchimento de falhas de acordo com as recomendações da Organização Mundial de Meteorologia; e (iv) geração de gráficos a partir dos dados armazenados (PRECINOTO *et al.*, 2013).

Resultados e Discussão

Esse trabalho teve início em maio de 2015 e, durante o período, foi contemplada a etapa de: criação de um *webcrawler*, como uma rotina que navega a internet de maneira sistemática e automatizada para carregar na base os dados de novas estações e dados meteorológicos disponibilizados. Dessa forma, aumentando a eficiência do processo de carga e pré-processamento dos dados.

A Figura 1 demonstra como o *webcrawler* se comunica com a plataforma Hidroweb para puxar os dados do servidor e carregar diretamente na base, ficando disponível assim para o pesquisador.

Os experimentos realizados utilizaram a base de dados histórica do sistema Hidroweb da Agência Nacional de Águas (ANA) contendo dados de precipitação pluvial, enquanto que as estações consideradas estão dentre as 555 localizadas no estado do Rio de Janeiro.

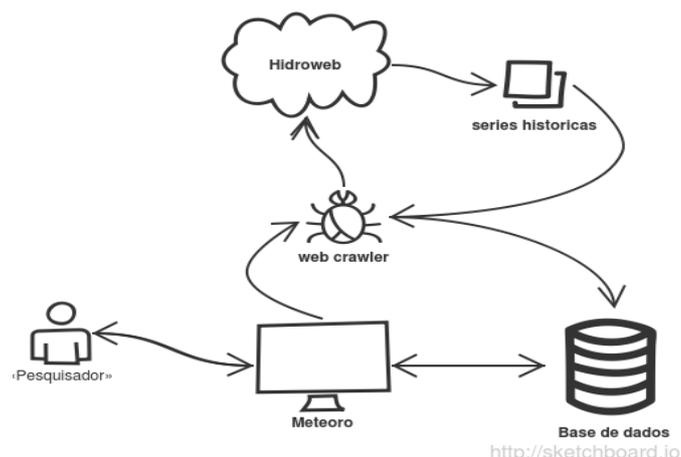


Figura 1 - Representação do *webcrawler* integrado ao sistema Meteoro.

Considerações Finais

Como está se manipulando grandes massas de dados na área de Meteorologia, requer novas abordagens, como por exemplo, técnicas de distribuição de dados e processos, mecanismos de coleta de descritores de proveniência e gerência semântica desses descritores, entre outros (CRUZ, 2011; CRUZ *et al.*, 2011, 2012). Sendo assim, serão implementados *workflows* científicos usando rotinas dos sistemas *VisTrails* (BAVOIL *et al.*, 2005) integrados ao sistema *Meteoro* para que se possa facilitar a organização e manipulação desses dados.

Em relação as próximas etapas espera-se: (i) ampliar os métodos de preenchimento de falhas e análise de consistência/homogeneidade para avaliação de dados meteorológicos, (ii) manter uma integração entre a plataforma *Meteoro* e os *workflows* desenvolvidos, (iii) analisar o controle de qualidade dos dados usando o sistema *Meteoro*, e (iv) expandir as variáveis meteorológicas contempladas pelo Sistema.

Agradecimentos

Agradecemos à FAPERJ pelo financiamento da Bolsa de Iniciação Científica que está tornando possível realizar este estudo.

Referências Bibliográficas

- BAVOIL, L.; CALLAHAN, S. P.; CROSSNO, P. J. *et al.*, 2005. [VisTrails: Enabling Interactive Multiple-View Visualizations](#). Proceedings of IEEE Visualization, pp. 135-142.
- CRUZ, S.M.S. *Uma Estratégia de Apoio à Gerência de Dados de Proveniência em Experimentos Científicos*. Tese de Doutorado. PESC/COPPE-UFRJ, 2011.
- CRUZ, S.M.S.; SILVA, C.E.P.; OLIVEIRA, D. *et al.*, 2011. Capturing Distributed Provenance Metadata from Cloud-Based Scientific Workflows. *J. Inform. and Data Management*, v. 2, n. 1, pp. 43-50.
- CRUZ, S.M.S.; MACHADO, M.L.M.; MATTOSO, M., 2012. A Foundational Ontology to Support Scientific Experiments. In: Proceedings of Joint V Seminar on Ontology Research in Brazil and VII International Workshop on Metamodels, Ontologies and Semantic Technologies, pp. 144-155, Recife, Pernambuco.
- HEY, T.; TANSLEY, S. & TOLLE, K., 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft, 287 p.
- LEMO FILHO, G. R.; PRECINOTO, R. S.; CORREIA, T. P.; SANTOS, E. O.; LYRA, G. B.; CRUZ, S. M. S., 2013. Assimilação, Controle de Qualidade e Análise de Dados de Meteorológicos Apoiados por Proveniência. XXXIII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO/BreSci – VII Brazilian e-Science Workshop, Maceió/AL, Julho/2013.
- VICENTE-SERRANO, S. M.; BERGUÉRIA-PORTUGUÉS, S. Estimating Extreme Dry-Spell Risk in the Middle Ebro Valley (Northeastern Spain): A Comparative Analysis of Partial Duration Series with a General Pareto Distribution and Annual Maxima Series With a Gumbel Distribution. *International Journal of Climatology*, 2003.
- VICENTE-SERRANO, S. M.; BERGUÉRIA-PORTUGUÉS, S.; LÓPEZ-MORENO, J. I. A Multi-scalar drought index sensitive to global warming: The Standardized Precipitation Evapotranspiration Index –SPEI. *Journal of Climate*, 23 (7), p. 1696-1718, 2010.
- WISSMANN, J.A.; TAMPELINI, L.G.; FEIL, A.C.; SAMPAIO, S.C.; SUSZEK, M. Ferramenta Computacional para Análise de Consistência de Dados Pluviométricos. *Varia Scientia*, Cascavel. Ago. 2006.