

Análise de tendências tecnológicas em recursos hídricos

Felipe Pedroni Ungaro¹

Maria Fernanda Moura²

A organização e análise da informação, neste trabalho, consistem em gerar uma metodologia e seu ferramental de software para analisar tendências em dados textuais que sejam mundialmente disponibilizados para acesso aberto, por meio de processos de mineração de textos. O uso de repositórios de acesso aberto é uma diretiva da Embrapa, que tem sido aplicada a seus próprios repositórios textuais. Essa iniciativa é bem representada pelo desenvolvimento e disponibilização do provedor de serviços Sistema Aberto e Integrado de Informação em Agricultura (SABIIA), que é o sistema responsável pela integração dos dados provenientes de repositórios institucionais, periódicos científicos, bibliotecas digitais e outros, tanto internos quanto externos, de interesse da Embrapa (VACARI et al., 2011). E, neste trabalho, os repositórios textuais de padrão de dados abertos são buscados de acordo com os temas de interesse de cada projeto de P&D, priorizando-se os que possuem citações sobre os temas de interesse.

Desta forma, o experimento aqui conduzido contribui com os objetivos do Projeto Componente de Tendências Tecnológicas da Rede AgroHIDRO - PC5. No PC5, pretende-se analisar as tendências tecnológicas, que contribuam para a preservação da qualidade da água, para seu uso eficiente, e para o aumento da produtividade da água na agricultura em bacias hidrográficas. Assim, os temas de interesse foram identificados pelos pesquisadores do PC5 e, a seguir, os repositórios foram selecionados de acordo com estes. Em termos de metodologias de mineração de textos, no experimento estão sendo avaliadas e integradas algumas ferramentas de mineração de textos e de busca, bem como de visualização de dados. A fim de realizar essa ava-

¹ Universidade Estadual de Campinas - felipe.ungaro@gmail.com

² Embrapa Informática Agropecuária - maria-fernanda.moura@embrapa.br

liação, na Figura 1, há uma representação esquemática de como ocorreram as buscas nos repositórios, e como as análises estão sendo processadas:

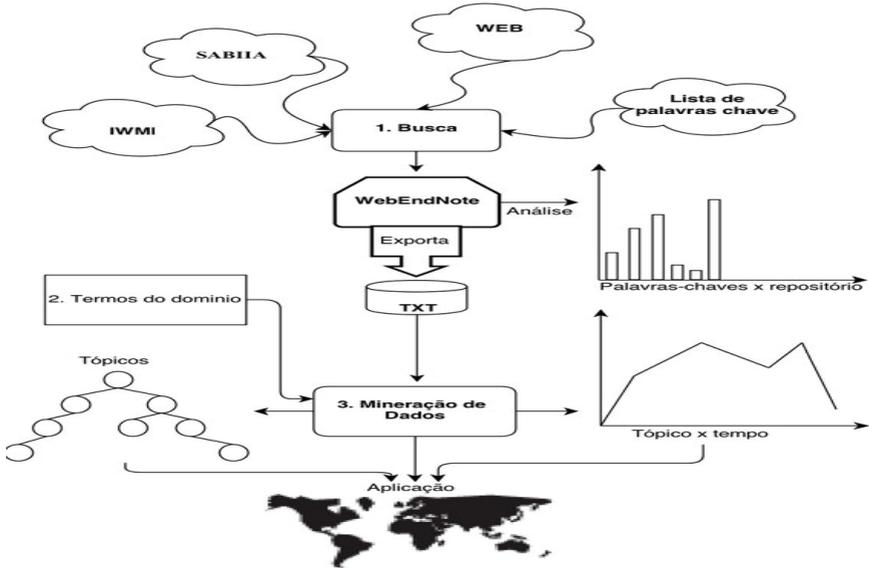


Figura 1. Fluxograma do processo.

- Etapa 1 - Busca: Foram realizadas buscas nos repositórios disponíveis pelo Sabiia, no site IWMI (<http://www.iwmi.cgiar.org/publications/library-catalog/>) e na WEB, de acordo com termos pré-definidos como relevantes pelos pesquisadores do PC5, por exemplo: balanço de energia, evapotranspiração, produção de biomassa, produtividade da água, sensoriamento remoto e geociências, em português, e *energy balance*, *evapotranspiration*, *biomass production*, *water productivity*, *remote sensing* e *geosciences*, em inglês. As buscas foram compiladas no WebEndNote e exportadas em formato que pudesse ser utilizado pelas ferramentas de mineração de textos. Ter utilizado o WebEndNote, possibilitou que a equipe do projeto pudesse acessar, validar e alterar os resultados das buscas. A partir das buscas realizadas também foi obtido um histograma de frequências, o qual indica os repositórios que apresentaram bons resultados.

- Etapa 2 - Foi construída uma lista de termos de domínio a partir de quatro glossários (JAMES, 1996; NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION, 2013; QUIZLET, 2013; UNESCO, 2013), os quais tiveram seus termos organizados em forma de planilha, alfabeticamente classificados e suas repetições eliminadas. Esta planilha será utilizada como vocabulário controlado para a realização do processo de mineração dos textos.
- Etapa 3 - Com o uso da planilha construída na Etapa 2 e dos textos da Etapa 1 foi iniciado o processo de extração e identificação de padrões, especialmente tópicos, por meio das ferramentas TORCH, TaxEdit e Mallet. Neste caso, os tópicos podem corresponder às tecnologias de interesse ou composições dessas com outras. Os tópicos resultantes do processo são também estudados de acordo com sua distribuição temporal e espacial.

Espera-se como resultado deste trabalho, delimitar uma metodologia, e ferramentas automatizadas, para identificar os tópicos encontrados nos textos, que correspondam às tendências de uso de tecnologias agrícolas, distribuídas no tempo e no espaço geográfico. E, com essa informação, complementar as informações obtidas no PC5 da Rede AgroHIDRO, de modo a auxiliar a análise de tendências tecnológicas que contribuam para a preservação da qualidade da água, seu uso eficiente e o aumento da produtividade da água na agricultura em bacias hidrográficas.

Referências

JAMES, W. **Hydraulics and hydrology vocabulary, Web Manual of Guelph, School of Engineering**. 1996. Disponível em: <http://www.soe.uoguelph.ca/webfiles/wjames/homepage/stillunclassified/wjvocab.html>>. Acesso em: 27 set. 2013.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION. **Glossary of hydrologic terms**. Disponível em: http://www.nws.noaa.gov/om/hod/SHManual/SHMan014_glossary.htm>. Acesso em: 27 set. 2013.

QUIZLET, **Hidrology vocabulary**. Disponível em: <http://quizlet.com/686409/hydrology-vocabulary-flash-cards/>>. Acesso em: 27 set. 2013.

UNESCO. **International glossary of hydrology**. Disponível em: <http://webworld.unesco.org/water/ihp/db/glossary/glu/HINDEN.HTM>>. Acesso em: 27 set. 2013.

VACARI, I.; VISOLI, M. C.; DEGAUT, S. D. de C. L.; OKAWACHI, M. F.; SIMÃO, V. P. M.; GONZALES, L. E.; PRAXEDES, M. G. G. Planejamento e desenvolvimento de serviço de informação em pesquisa agropecuária: caso do Sistema Aberto e Integrado de Informação em Agricultura (Sabiia). In: CONFERÊNCIA LUSO-BRASILEIRA ACESSO ABERTO, 2., 2011, Rio de Janeiro. **Anais...** Rio de Janeiro: IBICT, 2011. 3 p. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/48180/1/planejamento.pdf>>. Acesso em: 27 set. 2013.