

## SISTEMA DE RECUPERAÇÃO DE INFORMAÇÃO PARA PUBLICAÇÕES TÉCNICO-CIENTÍFICAS RELACIONADAS A RECURSOS HÍDRICOS

*Glauber José Vaz<sup>1\*</sup> & Leandro Henrique Mendonça de Oliveira<sup>2</sup> & Lineu Neiva Rodrigues<sup>3</sup>*

**Resumo** – A disseminação de informações confiáveis relacionadas a recursos hídricos é fundamental para a gestão da água no Brasil. Para isso, as ferramentas de busca são essenciais e precisam ser cada vez mais poderosas no cenário atual de transformação digital, em que a experiência dos usuários ocupa lugar central diante das profundas mudanças provocadas nas organizações pelas tecnologias digitais. O objetivo deste trabalho foi desenvolver um sistema de recuperação de informação especializado em publicações técnico-científicas da área de recursos hídricos, que inclui recursos de glossário e de visualização de elementos geográficos em mapa. Baseado em APIs e em componentes visuais, o sistema pode agregar novos recursos de maneira simples, abrindo caminho para inúmeras oportunidades de evolução. O sistema possibilita a recuperação de informações presentes em publicações técnico-científicas relacionadas a recursos hídricos, contribuindo para a disseminação de informações confiáveis que são fundamentais para a tomada de decisões seguras e responsáveis na gestão de recursos hídricos.

**Palavras-Chave** – Sistema de recuperação de informação; Planejamento e gestão de recursos hídricos; Gestão de conhecimento técnico-científico.

## INFORMATION RETRIEVAL SYSTEM FOR TECHNICAL AND SCIENTIFIC PUBLICATIONS RELATED TO WATER RESOURCES

**Abstract** – The dissemination of reliable information related to water resources is fundamental for water management in Brazil. For this, search tools are essential and need to be increasingly powerful in the current scenario of digital transformation, where the user experience occupies a central place in the face of the profound changes brought about by digital technologies in organizations. The objective of this work was to develop an information retrieval system specialized in technical and scientific publications in the area of water resources. Based on APIs and visual components, the system can aggregate new features in a simple way, paving the way for numerous development opportunities. It enables the retrieval of information in technical and scientific publications related to water resources, contributing to the dissemination of reliable information, which is fundamental for making safe and responsible decisions in the management of water resources.

**Keywords** – Information retrieval system; Water resources planning and management; Management of technical and scientific knowledge.

<sup>1</sup> Embrapa Informática Agropecuária, glauber.vaz@embrapa.br.

<sup>2</sup> Embrapa. Departamento de Pesquisa e Desenvolvimento. Coordenadoria de Gestão da Informação, leandro.oliveira@embrapa.br.

<sup>3</sup> Embrapa Cerrados, lineu.rodrigues@embrapa.br.

## INTRODUÇÃO

A água é um recurso vital para o desenvolvimento de qualquer atividade econômica, mas a sua disponibilidade é limitada e variável de ano para ano. A sua gestão deve sempre proporcionar o uso múltiplo das águas, preconizada na visão estratégica do estudo de planejamento territorial, que requer uma atuação integrada dos diversos setores que concorrem pelo seu uso: consumo humano, produção e energia.

Os instrumentos da Política Nacional de Recursos Hídricos (PNRH), em especial os Planos de Recursos Hídricos e Sistema de Informações sobre Recursos Hídricos, dependem, para sua concepção e operacionalização, de uma boa base de informações, que nem sempre estão facilmente acessíveis. Além disso, segundo o Plano Nacional de Recursos Hídricos (BRASIL, 2006), a disseminação de informações confiáveis é peça fundamental para a tomada de decisões seguras e responsáveis por parte das comunidades, dos usuários de água e do poder público. O Plano ainda aponta o fortalecimento da política de capacitação em Ciência e Tecnologia como elemento importante na construção de uma estratégia integrada dos recursos hídricos do Brasil, envolvendo inclusive a capacitação informal, aberta a toda a sociedade, por meio da disseminação de informações que tratem dos aspectos relacionados à água e ao meio ambiente.

Melhorar a eficiência e a eficácia da recuperação de informações técnico-científicas relacionadas ao tema recursos hídricos, por meio de ferramentas mais adequadas, pode contribuir para fortalecer a governança, o planejamento, a gestão e os instrumentos da PNRH. Para isso, é necessário incorporar e utilizar cada vez mais na área de recursos hídricos técnicas e ferramentas já disponíveis e utilizadas pelas ciências da computação e da informação.

À medida que os usuários se acostumam a realizar buscas e a utilizar a Internet, eles começam a demandar ferramentas mais poderosas para suas necessidades, de maneira que, em buscas por documentos na Internet, o foco deve mudar dos documentos que estão sendo pesquisados para as tarefas que as pessoas precisam realizar (HOEBER, 2008).

Essa maior atenção ao papel do usuário e o oferecimento de uma melhor experiência digital, em que conteúdo e funcionalidades são adaptados às necessidades dos usuários, estão associados ao que vem sendo chamado de transformação digital, que exige das organizações mudanças profundas na forma de se fazer negócios devido às tecnologias digitais (EARLEY, 2014; BASOLE, 2016). As APIs (*Application Programming Interface*), que especificam como os componentes de software podem interagir entre si, exercem papel fundamental nesse contexto facilitando a integração de sistemas e dispositivos, inclusive externos.

No cenário atual em que há uma grande disponibilidade de APIs, as de busca ocupam lugar de destaque, pois possibilitam o acesso rápido a informações relevantes e são consideradas fundamentais para serviços digitais (BASOLE et al., 2016).

No entanto, as APIs são voltadas a desenvolvedores que podem criar novas soluções com os serviços disponibilizados. Para se alcançar efetivamente os clientes finais dos sistemas de informação, é necessário oferecer interface de usuário que possibilite uma riqueza na experiência de uso. Normalmente, as pessoas usam estratégias de busca baseando-se nas funcionalidades oferecidas pelo sistema. Assim, interfaces que oferecem métodos mais flexíveis, combinando múltiplas técnicas em uma interface única, por exemplo, viabilizam o uso de diferentes estratégias de busca por informação e de interação com a interface (RUTHVEN, 2008).

Também pode ser determinante na experiência do usuário o grau de especialização das ferramentas. Aplicações genéricas são capazes de atender a um público amplo, mas soluções mais específicas normalmente são necessárias para satisfazer demandas particulares de um determinado grupo de usuários.

Embora a recuperação de informação seja de fundamental importância na área de recursos hídricos, se desconhece uma ferramenta de busca de informações técnico-científicas que explore o conhecimento especializado sobre os recursos hídricos do país.

Além disso, o oferecimento de uma interface que promova uma melhor experiência digital ao usuário e a disponibilização de uma API para que terceiros possam criar novas soluções baseadas nos dados das publicações consideradas são de grande importância para ampliar o alcance das ferramentas de busca.

Desta forma, o presente trabalho teve como objetivo desenvolver um sistema de recuperação de informações técnico-científicas especializado em recursos hídricos, visando principalmente a disseminação de informações confiáveis no tema.

## **METODOLOGIA**

O sistema de recuperação de informação desenvolvido nesse trabalho foi construído utilizando-se basicamente três principais tecnologias: uma ferramenta de indexação e busca de documentos que oferece acesso via API, um gerenciador de APIs e ainda uma ferramenta de portal baseada em componentes visuais (portlets).

O Liferay Portal, considerada uma das líderes no segmento de portais pelo Gartner Inc (MURPHY et. al, 2016), foi a ferramenta utilizada. O sistema de recuperação de informação é disponibilizado em uma única página que conta com vários portlets, cada um responsável por uma funcionalidade específica.

Para o servidor de busca foi utilizado o Elasticsearch, tecnologia de código aberto que oferece interface simples via APIs (GORMLEY, C.; TONG, 2015), possibilitando, assim, que diferentes aplicações possam acessar os recursos disponibilizados pelo servidor. Elasticsearch possibilita que várias fontes de dados sejam utilizadas para se criar os índices e que, além de buscas textuais, sejam realizadas buscas estruturadas e análises de dados.

Para o gerenciamento da API de busca foi utilizada a ferramenta da WSO2, também de código aberto e que oferece as funcionalidades necessárias para o adequado gerenciamento das APIs, como criação, publicação, gerenciamento do ciclo de vida, versionamento e governança, além de interface para implantação e monitoramento das APIs e para a inscrição dos consumidores (WSO2 INC., 2017).

As publicações utilizadas são as que constam no repositório de documentos da Empresa Brasileira de Pesquisa Agropecuária (Embrapa). Assim, documentos sobre diferentes temas podem ser recuperados nas consultas realizadas no sistema, mas os resultados para pesquisas relacionadas a recursos hídricos apresentam maior detalhamento de informações.

Em relação às funcionalidades do sistema, foram considerados neste trabalho a visualização de um glossário sobre recursos hídricos e o escopo geográfico de documentos, que é o conjunto de lugares e regiões associados ao conteúdo de um documento.

Descobrir lugares relacionados ao conteúdo de um documento textual, conhecido como o problema de resolução de escopo geográfico, ainda apresenta inúmeros desafios de pesquisa. Embora avanços importantes estejam sendo alcançados na área, questões de ambiguidades e incertezas, por exemplo, ainda precisam avançar bastante (MONTEIRO et. al., 2016). Dado o estado da arte na determinação do escopo geográfico de forma automática, foi considerado que essa informação deve ser catalogada por um profissional para que o sistema funcione adequadamente. Instituições que contam com profissionais dedicados a executar processos bibliográficos bem estabelecidos não precisam de grandes alterações em seus processos para possibilitar o oferecimento deste recurso aos usuários de suas publicações.

A exibição de um glossário, em que são fornecidos os termos de uma área do conhecimento, seus sinônimos e suas definições, auxiliam na consolidação de um entendimento comum da terminologia usada pelos usuários do sistema. Também agiliza a verificação do significado dos termos presentes nas publicações consultadas sem a necessidade de recorrer a outras ferramentas. Finalmente, o glossário pode ajudar na estratégia de busca do usuário ao exibir nas definições dos termos palavras que são importantes para a sua consulta.

## RESULTADOS E DISCUSSÃO

A Figura 1 exibe a interface do sistema de recuperação de informação criado com quatro portlets: (i) Formulário: responsável por exibir o formulário de entrada para a realização da busca; (ii) Resultados: posicionado à direita na Figura 1, exibe dados dos documentos que atendem a uma pesquisa de usuário ou as informações de um documento específico quando este é selecionado. Este portlet faz a paginação dos resultados, exibindo no máximo dez itens por página, valor este que pode ser configurado; (iii) Glossário: mostra a definição de termos presentes nas publicações descritas no componente Resultados; (iv) Mapa: exibe em um mapa os elementos geográficos associados aos documentos listados em Resultados.

As informações exibidas nos portlets são consultadas no servidor de busca por meio da API disponibilizada. Em princípio, para atender às demandas destes componentes, quatro tipos de chamadas ao servidor são possíveis, todas elas envolvendo passagem de parâmetros: (i) Obtenção dos dados de um determinado documento a partir de seu identificador. Usada para exibir os dados de um documento que foi selecionado; (ii) Obtenção da lista de documentos a partir de uma busca realizada. Neste caso, além da consulta submetida pelo usuário, também são enviados os parâmetros que possibilitam a paginação dos resultados: a quantidade de itens que devem ser exibidos por página e a posição na lista de resultados do primeiro item que deve ser exibido na página corrente; (iii) Obtenção dos termos do glossário presentes em um determinado documento, cujo identificador é fornecido. No caso em que um documento é selecionado, apenas uma chamada é suficiente para se obter os termos deste documento, mas se o portlet Resultados exibir vários itens, uma chamada para cada item é necessária a fim de se obter todos os termos presentes nos documentos listados; (iv) Obtenção de definições para termos consultados. Retorna a lista de definições para um ou mais termos enviados por parâmetro.

O glossário utilizado no sistema conta atualmente com 86 termos relacionados a recursos hídricos. Para cada termo, há uma definição estabelecida por uma equipe que envolveu especialistas da área e linguistas. Já a utilização de elementos geográficos ainda precisa ser incorporada nos processos da organização. Os elementos considerados aqui, embora reais, foram utilizados para teste de conceito. O sistema está pronto para utilizar dados geográficos, mas é necessário que a

catalogação das publicações possibilite a inclusão de dados geográficos em formato GeoJSON, quando se deseja fazer esta associação entre publicações e elementos geográficos.

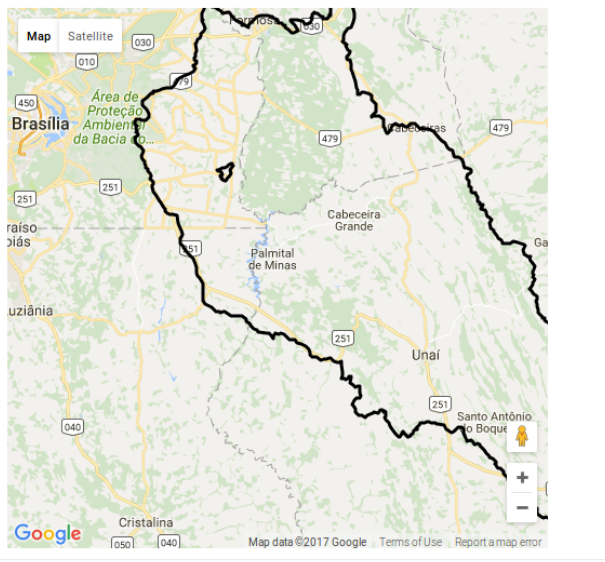
### Formulário

### Glossário

**Bacia hidrográfica:** área onde ocorre a captação natural das águas das chuvas para um rio principal e seus afluentes, promovida pelo desnível dos terrenos, que direciona os cursos da água das áreas mais altas para as mais baixas.

**Uso da terra:** Modo como os elementos biofísicos da superfície da Terra são utilizados pelo homem.

### Mapa



### Resultados

Documentos de 1 a 10. Total: 17

**CAMPOS, J. E. G.; MONTEIRO, C. F.; RODRIGUES, L. N.** [Geologia e zoneamento hidrogeológico da bacia do Rio Preto, DF/GO/MG](#). Planaltina, DF: Embrapa Cerrados, 2006. 2006

Bacia hidrográfica do Rio Preto  
**Palavras-chave:** Bacia hidrográfica; Geologia; Geology; Recurso hídrico; Water resources; Watersheds; Zoneamento; Zoning

**AMARAL, L. S.; SILVA, L. J. L.; CORREIA, J. R.; SILVA, E. M.; RODRIGUES, L. N.; BISPO, J. P. C.** [Mapeamento pedológico da bacia hidrográfica do Rio Buriti Vermelho, Distrito Federal](#). In: ENCONTRO DE JOVENS TALENTOS DA EMBRAPA CERRADOS, 3., 2007, Planaltina, DF. Resumos apresentados. Planaltina, DF: Embrapa Cerrados, 2007. p. 42. 2007

Bacia experimental do Rio Buriti Vermelho  
**Palavras-chave:** Mapa; Solo

**CASTRO, K. B. DE; MARTINS, E. de S.; BRAGA, A. R. dos S.; LIMA, L. A. DOS S.; RODRIGUES, L. N.** [Manejo da terra e compartimentação geomorfológica da Bacia Hidrográfica do Rio Buriti Vermelho, DF](#). In: ENCONTRO DE JOVENS TALENTOS DA EMBRAPA CERRADOS, 4., 2009, Planaltina, DF. Resumos apresentados... Planaltina, DF: Embrapa Cerrados, 2009. 2009

**Palavras-chave:** Geoprocessamento

**MANETA, M. P.; SINGH, P. N.; TORRES, M.; WALLENDER, W. W.; VOSTI, S. A.; RODRIGUES, L. N.; BASSOI, L. H.; YOUNG, J. A.** [A parsimonious crop-water productivity index: an application to Brazil](#). Area, v. 41, n. 1, p. 94-106, 2009. 2009

**Palavras-chave:** Bacia hidrográfica; Desenvolvimento agrícola; Irrigação; Irrigation; Manejo de água; Produtividade agrícola

**PASSO, D. P.; CARDOSO, W. dos S.; MARTINS, E. de S.; BRAGA, A. R. dos S.; RODRIGUES, L. N.; SANTANA, O. A.; CARVALHO JÚNIOR, O. A. de.; GOMES, R. A. T.; LIMA, L. A. de S.; CASTRO, K. B. de.; SAMPAIO, D. R.; SANTOS, L. L. de B. dos.** [Influência da textura e do tipo de uso do solo na condutividade hidráulica saturada da Bacia Hidrográfica do Rio Buriti Vermelho, DF](#). In: ENCONTRO DE JOVENS TALENTOS DA EMBRAPA CERRADOS, 4., 2009, Planaltina, DF. Resumos apresentados... Planaltina, DF: Embrapa Cerrados, 2009. 2009

**Palavras-chave:** Conservação; Manejo; Solos

**CASTRO, K. B. de; MARTINS, E. de S.; BRAGA, A. R. dos S.; LIMA, L. A. de S.; RODRIGUES, L. N.; CARVALHO JUNIOR, O. A.; BORGES, M. E. S.; SOUZA, V. V. de; GOMES, M. P.** [Compartimentação geomorfológica da bacia hidrográfica do rio Buriti Vermelho, Distrito Federal, DF](#). Planaltina, DF: Embrapa Cerrados, 2009. 2009

**Palavras-chave:** Geoprocessamento; Geoprocessing; Land use; Soil; Solo; Uso da terra

Figura 1 - Interface de usuário do sistema de recuperação de informação

Assim como novos portlets podem ser criados e agregados à interface de usuário de maneira simples, novas funcionalidades podem ser adicionadas à API. No entanto, a instituição que oferece os recursos precisa estar preparada para disponibilizar com qualidade os dados e os serviços necessários para o sistema. No caso de georreferenciamento das publicações, por exemplo, deve haver sistemas de gerenciamento de publicações adequados e processos internos que viabilizem a disponibilização destas informações com o trabalho de profissionais habilitados. Da mesma forma, para se tirar proveito de um glossário conforme explicado, é necessário que profissionais especializados criem as definições dos termos de maneira adequada e façam a validação junto a especialistas do assunto para que as informações utilizadas sejam confiáveis. Além disso, esse conhecimento deve ser disponibilizado de maneira que possa ser manipulado por computadores.

Embora as funcionalidades atuais do sistema tenham sido apresentadas, sua característica mais importante é a capacidade de se expandir para agregar novas funcionalidades e possibilidades de interação, além de oferecer oportunidades para a criação de novas soluções por parte de desenvolvedores que acessem a API disponibilizada. Exemplos de recursos avançados que podem ser incorporados são aqueles que envolvem técnicas de inteligência artificial, como os componentes desenvolvidos para o CiteSeerX (WU et al, 2014).

## CONCLUSÕES

O sistema apresentado possibilita a recuperação de informações de publicações técnico-científicas relacionadas a recursos hídricos, contribuindo para a disseminação de informações e para a gestão da água. Baseado em portlets, o sistema integra múltiplas funcionalidades em uma interface única, de maneira que possibilita o uso de diferentes estratégias de busca e de interação e proporciona uma melhor experiência ao usuário. Além disso, oferece APIs que viabilizam o desenvolvimento de novas soluções por terceiros. Embora o sistema ainda apresente poucos recursos, como a exibição de glossário do domínio e de elementos geográficos em mapa, está estruturado para integrar novas funcionalidades de maneira muito simples.

Este trabalho pode ter inúmeras frentes de evolução. Conhecimentos de diferentes domínios podem ser incorporados, assim como outras bases de publicações. Visualização de informações relacionadas às publicações e geração de dados a partir de técnicas de inteligência artificial abrem vasto leque de oportunidades. Várias formas de interação, como, por exemplo, avaliação de publicações pelos usuários e comentários sobre documentos ou definições de termos também são possibilidades.

## REFERÊNCIAS

- BASOLE, R. C. (2016). Accelerating Digital Transformation: Visual Insights from the API Ecosystem. *IT Professional*, 18(6), pp. 20-25.
- BRASIL. Secretaria de Recursos Hídricos. Plano Nacional de Recursos Hídricos: síntese executiva. Brasília, DF, 2006. 135 p.
- EARLEY, S. (2014). The Digital Transformation: Staying Competitive. *IT Professional*, 16(2), pp. 58-60.
- GORMLEY, C.; TONG, Z. (2015) Elasticsearch: the definitive guide. Disponível em: <https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html>. Acesso em: 3 mai. 2017.
- HOEBER, O. (2008). Web information retrieval support systems: The future of web search. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 03*, Dez. 2008, pp. 29-32. IEEE Computer Society.
- MONTEIRO, B. R.; DAVIS, C. A.; FONSECA, F. (2016). A survey on the geographic scope of textual documents. *Computers & Geosciences*, 96, 23-34.
- MURPHY, J.; PHIFER, G; TAY, G; REVANG, M. (2016) Magic quadrant for horizontal portals. Out. 2016.
- RUTHVEN, I. (2008). Interactive information retrieval. *Annual review of information science and technology*, 42(1), pp. 43-91.

WSO2 INC. (2017) WSO2 API Manager documentation: introducing the API Manager. Disponível em <https://docs.wso2.com/display/AM210/WSO2+API+Manager+Documentation>. Acesso em 3 mai. 2017.

WU, J. et al. (2014). CiteSeerX: AI in a Digital Library Search Engine. In *AAAI*. Jul. 2014, pp. 2930-2937.