

Sistemas de Recuperação da Informação em Ambientes Informacionais Digitais no Contexto de Incêndios Florestais

Gustavo Marttos Cáceres Pereira

Programa de Pós-Graduação em Ciência da Informação

Universidade Estadual Paulista "Júlio de Mesquita Filho", Marília, São Paulo, Brasil

gustavo.marttos@unesp.br

Abstract—The information retrieval is present in all informational environments, especially with regard to digital informational environments, where these are covered with an information retrieval system. The storage, indexing and document representation are important factors that impact directly on information recovery, regardless of the environment. In the case of news portals, informational environments that have digital documents on various subjects, such as forest fires, need to have implemented the information retrieval systems able to meet the needs of users.

Index Terms—informational retrieval, digital informational environments, forest fires

Resumo—A recuperação da informação está presente em todos os ambientes informacionais, principalmente no que se refere aos ambientes informacionais digitais, onde estes são contemplados com um sistema de recuperação da informação. O armazenamento, indexação e representação de documentos são fatores importantes que impactam diretamente na recuperação da informação, independente do ambiente. No caso de portais de notícias, ambientes informacionais digitais que possuem documentos sobre diversos assuntos, como por exemplo de incêndios florestais, precisam ter implementado sistemas de recuperação da informação capazes de atender às necessidades dos usuários.

Palavras-chave—recuperação da informação, ambientes informacionais digitais, incêndios florestais

I. INTRODUÇÃO

A Recuperação da Informação é uma área fundamentada na Ciência da Computação, cujo termo foi cunhado por [1] como uma área capaz de englobar aspectos intelectuais tanto da descrição de informações, quanto de suas especificidades para a busca, além de quaisquer sistemas, técnicas ou máquinas empregados para o desempenho da operação.

A partir disso, a recuperação da informação passa a ser considerada como uma solução bem sucedida na Ciência da Informação [2]. De acordo com a literatura, [1] identifica três problemas fundamentais e pertinentes na área: (1) como descrever intelectualmente a informação?; (2) como especificar intelectualmente a busca?; e (3) que sistemas, técnicas ou máquinas devem ser empregados?

Para responder essas perguntas, [2] resume que as pesquisas exploratórias na área de RI, as quais levantaram os fenômenos, processos e variáveis, bem como as causas, efeitos, comportamentos e manifestações, foram responsáveis pelo desenvolvimento de inúmeras aplicações, como produtos, sistemas, redes e demais serviços.

Ademais, um sistema digital de recuperação da informação passa a integrar um ambiente informacional digital, sendo este definido por [3] como um local que reúne informações a fim de minimizar as necessidades informacionais. É válido ressaltar que os sistemas de recuperação da informação inseridos em ambientes informacionais digitais estão sob a perspectiva do usuário, afinal, como dito anteriormente, o objetivo é atender suas necessidades informacionais.

O início dessa pesquisa teve como gatilho um comum problema no bioma Cerrado do Distrito Federal (DF), o qual atualmente apresenta um cenário de recorrência e intensificação de incêndios florestais. As áreas protegidas no Cerrado, relativamente isoladas entre áreas urbanizadas, são afetadas anualmente, conforme demonstrado pelo relatório do Instituto Brasília Ambiental [4], apesar da existência de um Plano local para a Prevenção e Combate a Incêndios Florestais [5].

No contexto de incêndios florestais, os quais são relatados por meio do serviço público de emergências 193, também podem ser constatados em outros meios digitais, como portais de notícias. Tais portais são considerados como ambientes informacionais digitais e proveem de sistemas digitais de recuperação da informação.

As informações recuperadas podem ser úteis e relevantes para auxiliar os usuários, no caso os operadores de um ambiente emergencial digital, pois complementam as informações prévias que estes possuem, melhorando assim sua qualidade. Por outro lado, quando um sistema de recuperação de informação não é bem elaborado e não destaca a importância da necessidade do usuário, as informações recuperadas podem ser prejudiciais ou não ser relevantes por não contribuírem para com os usuários.

Visto isto, se faz necessário planejar, desenvolver e analisar criteriosamente os sistemas digitais de recuperação da informação a fim de minimizar as necessidades informacionais dos usuários, principalmente em portais de notícias, estes sendo considerados como ambientes informacionais digitais.

Como resultado, esta pesquisa traz uma análise e uma discussão a respeito de de um determinado portal de notícias do Brasil, bem como a importância dos sistemas de recuperação da informação nos ambientes informacionais digitais de modo que supram as necessidades daqueles que os utilizam.

II. SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO

A definição de sistemas de recuperação da informação é encontrada tanto na Ciência da Informação, quanto na Ciência da Computação, permitindo a interação entre ambas as disciplinas, convergindo seus conceitos, técnicas e métodos. Devido a este fato, [6] tem no espectro da informação o sentido de coisa, ou seja, o conhecimento registrado em documentos. Os documentos, por sua vez, de acordo com [7], são quaisquer signos físicos ou simbólicos, preservados ou registrados com o objetivo de representar, reconstruir ou demonstrar fenômenos físicos ou abstratos.

[8] assume que os sistemas de recuperação da informação sejam capazes de organizar e viabilizar o acesso aos documentos desde que possuam funcionalidades que desempenhem atividades de: (1) retratar as informações inclusas nos documentos utilizando processos de indexação; (2) armazenar os documentos, bem como suas respectivas representações; e (3) representar as informações recuperadas de acordo com a satisfação da necessidade do usuário.

A partir dos pontos levantados por [8], nota-se que a representação é um fator crucial, porém sabe-se que para esta gerar bons resultados é necessário que haja uma boa indexação dos documentos armazenados, sendo esta a principal função em sistema de recuperação da informação [9].

No que condiz propriamente dito aos sistemas de recuperação da informação, [10] apresentam dois modelos: (1) clássico, constituído pelos modelos booleano, vetorial e probabilístico; e (2) modelos mais avançados, sendo estes o modelo fuzzy, redes neurais e algoritmos genéticos.

Tais modelos, considerando que os mais avançados foram elaborados a partir do clássico, permitiram maior operabilidade e melhoria da performance, afirmando que a importância da indexação, bem como dito por [9].

A indexação, por sua vez, objetiva representar o seu conteúdo temático por meio de um conjunto de termos a fim de sumarizar seu conteúdo, ressaltando aquilo que o torna essencial. Ademais, eles servem de subsídio para localizar documentos a serem recuperados por um sistema de recuperação da informação.

[11] pontua dois tipos de indexação: (1) indexação por extração, considerando termos específicos do próprio documento, utilizando normativas pessoais e institucionais para selecionar quais termos representarão aquele documento; e (2) indexação por atribuição, utilizando um vocabulário (tesauro ou ontologia) com termos previamente definidos e normalizados.

Sistemas de recuperação de informação devem possuir uma boa indexação, afinal isto proporciona uma melhor representação de resultados dos documentos frente à necessidade do usuário.

III. AMBIENTES INFORMACIONAIS DIGITAIS

Na Ciência da Informação, é mais apropriado dizer que o termo ambiente refere-se aos contextos informacionais, uma vez que neste abrange todo o potencial de armazenamento de informações, além de outros sujeitos, como recursos humanos – ou usuários, como já dito anteriormente. Para

[12], a percepção do ambiente proporciona aos usuários a recuperação, representação, acesso, uso e modificação de informações.

[13] já considerava em organizar e representar informações em modelos arquitetônicos em diversos meios, como mapas, guias e materiais impressos, possibilitando futuramente sua recuperação por humanos. No que se refere à Recuperação da Informação, as ideias de [14] quanto aos ambientes informacionais foram pensadas para os usuários, ou seja, para que haja a interação usuário-informação, podendo assim encontrar aqui que de fato precisem, sanando suas necessidades de informação.

Os autores ainda ressaltam que dentro de um ambiente informacional digital, o sistema de busca possibilita a recuperação da informação para que o usuário possa utilizá-la de acordo com suas demandas cognitivas [14]. É válido ressaltar que tanto a recuperação quanto à representação são fatores impactantes num mecanismo de busca, pois a primeira deverá retornar ao usuário aquilo que este desejou, enquanto que na representação deve conter elementos como vocabulários controlados, metadados e outros elementos que facilitem a descrição do documento a fim de promover a compreensão do usuário.

A. Portais de notícias como ambientes informacionais digitais

[12] argumentam que um ambiente informacional digital é composto por sistemas de navegação, rotulagem, busca (ou recuperação) e representação, permitindo que o usuário interaja com o ambiente.

Partindo dessa premissa, portais de notícias são considerados ambientes informacionais digitais por atenderem à composição dos sistemas abordados na arquitetura informacional.

Os documentos armazenados em portais de notícias abrangem diversas temáticas, considerando a relevância dos assuntos de interesse por parte do público alvo. Grandes portais de notícias possuem vocabulários controlados, como os tesouros, a fim de proporcionar melhores indexações e representações.

Por estarem diretamente frente aos interesses dos usuários, os portais de notícias possuem documentos armazenados cujas informações são voltadas para a sociedade como um todo, partindo desde aspectos sociais, econômicos, políticos, ambientais, segurança etc. Tais documentos podem ser recuperados por meio do sistema de recuperação da informação desses portais, utilizando mecanismos de buscas que interajam diretamente com o usuário, como buscar por termos específicos ou palavras-chave, por títulos completos ou até mesmo por trechos da descrição do documento.

No que se refere a incêndios florestais, além do serviço público de emergências 193, os portais de notícias também são utilizados para verificar ou alertar sobre casos de incêndios, uma vez que os portais classificam essas ocorrências vinculadas às questões ambientais.

IV. ANÁLISE E DISCUSSÃO SOBRE OS PORTAIS DE NOTÍCIAS DO DISTRITO FEDERAL

Com o objetivo de analisar e discutir a respeito da recuperação da informação em ambientes informacionais digitais, o portal de notícias Correio Braziliense¹ é considerado o mais abrangente na região do Distrito Federal e, portanto, foi selecionado para a análise de seu sistema de recuperação de informação no contexto de incêndios florestais.

A primeira vista, o mecanismo de busca do portal de notícias selecionado parece bastante simples, sendo composto apenas por um campo de digitação e um botão de envio para acionar a funcionalidade específica de busca, conforme apresentado na Figura 1.

Como usuário, a necessidade informacional é representada pelo termo “incêndio florestal” para busca e futuramente sua recuperação. Ao efetivar essa ação no mecanismo de busca, conforme apresentado na Figura 2, alguns resultados são apresentados, sendo estes considerados como documentos, conforme Figura 3.

Os principais documentos recuperados, destacados na Figura 3, demonstram três situações de incêndios florestais, porém nenhuma que se refira à localização do Distrito Federal. Apenas casos de incêndios internacionais, como as ocorrências da Califórnia, Portugal e Canadá.

A partir dessa breve análise e considerando o perfil do usuário como um sujeito que dependa de informações sobre incêndios florestais no Distrito Federal, os documentos recuperados seriam irrelevantes. Entretanto, nota-se que o sistema de recuperação da informação trouxe exatamente aquilo que estava de acordo com a necessidade do usuário, ou seja, recuperar documentos que contenham informações sobre incêndios florestais.

O mecanismo de busca ainda conta com recursos mais aprimorados, porém que pertencem ao modelo booleano de recuperação da informação, disponibilizando quais seções devem ser consideradas como filtros para a recuperação informacional, conforme é apresentado na Figura 4.

Ademais, é possível notar que os documentos podem ser ordenados por ordem cronológica ou por relevância, entretanto não se sabe qual é o critério de relevância utilizado.

V. CONCLUSÃO

Apesar do objetivo de busca não ter sido atingido, ou seja, recuperar informações em documentos que contenham termos relacionados a incêndios florestais utilizando como fator intrínseco a localização do portal de notícias no Distrito Federal, foi constatado que o sistema de recuperação de informação desse ambiente informacional digital conta com os requisitos necessários listados por [14] para se obter resultados de acordo com a necessidade do usuário.

A indexação é dada como fator de importância, justamente para que o sistema de recuperação da informação seja capaz de classificar e retornar os documentos esperados. Ademais, assim como apontado por [8], o sistema de recuperação de informação possui tanto a capacidade de armazenamento de

documentos, bem como a de suas representações, como visto nas figuras anteriores.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- [1] C. N. Mooers, “Zatocoding applied to mechanical organization of knowledge,” *American documentation*, vol. 2, no. 1, pp. 20–32, 1951.
- [2] T. Saracevic, “Ciência da informação: origem, evolução e relações,” *Perspectivas em ciência da informação*, vol. 1, no. 1, pp. 41–62, 1996.
- [3] L. S. d. A. d. Camargo, “Metodologia de desenvolvimento de ambientes informacionais digitais a partir dos princípios da arquitetura da informação,” Ph.D. dissertation, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), 2010.
- [4] G. do Distrito Federal, “Relatório de Área queimada nos parques e unidades de conservação do distrito federal do ano de 2015.” Disponível em http://www.ibram.df.gov.br/wp-content/uploads/2018/02/Promaq_Relat%C3%B3rios-de-%C3%A1reas-queimadas.compressed.pdf. Acessado em 11 ago. 2018, 2016.
- [5] —, “Decreto no. 37.549 - plano de ação de prevenção e combate aos incêndios florestais do distrito federal - ppcif,” Disponível em http://www.sinj.df.gov.br/sinj/Norma/22f0bba5f2b543bfb4038ba133f5ac5a/exec_dec_37549_2016.html. Acessado em 11 ago. 2018, 2016.
- [6] M. K. Buckland, “Information as thing,” *Journal of the American Society for information science*, vol. 42, no. 5, pp. 351–360, 1991.
- [7] S. Briet and L. Martinet, *What is documentation?: English translation of the classic French text*. Scarecrow Press, 2006.
- [8] R. R. Souza, “Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências,” *Perspectivas em ciência da informação*, vol. 11, no. 2, pp. 161–173, 2006.
- [9] V. M. A. P. d. Araújo, “Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos,” *Transinformação*, vol. 24, no. 2, 2012.
- [10] R. E. d. Silva, P. L. V. A. d. C. Santos, E. Ferneda, *et al.*, “Modelos de recuperação de informação e web semântica: a questão da relevância,” *Informação & Informação*, vol. 18, no. 3, pp. 27–44, 2013.
- [11] F. W. Lancaster, “Indexação e resumos: teoria e prática. tradução de antônio agenor briquet de lemos. rev. atual.” 2004.
- [12] H. P. C. d. Oliveira and S. A. B. G. Vidotti, “Dos ambientes informacionais às ecologias informacionais complexas,” *Informação & Sociedade: Estudos*, vol. 26, no. 1, 2016.
- [13] R. S. Wurman, “Information architect,” *Information Architects*, 1996.
- [14] P. Morville and L. Rosenfeld, *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*. O’Reilly Media, Inc., 2006.

¹Veja <https://www.correiobraziliense.com.br/>. Acesso em 17 de dezembro de 2018.

CORREIO BRAZILIENSE

Fig. 1. Formulário de busca do Correio Braziliense

10.275 resultados para
"incêndio florestal"



27/7 - Incêndio florestal na Califórnia

10:40 - 27/07/2018 - Compartilhe [f](#) [t](#)

Incêndios florestais continuam em Portugal

"Vários habitantes tiveram que ser tratados pelos médicos com problemas respiratórios, o fogo se aproximou de suas casas. Nos fazem falta meios para combater o incêndio", declarou o porta-voz dos bombeiros

17:49 - 06/09/2016 - Compartilhe [f](#) [t](#)

Incêndios florestais atingem Canadá

Assim como o estado americano da Califórnia, a Colúmbia Britânica é afetada por incêndios com frequência

19:47 - 18/07/2017 - Compartilhe [f](#) [t](#)

Fig. 2. Termo "incêndio florestal" aplicado no mecanismo de busca

CORREIO BRAZILIENSE

Fig. 3. Documentos (notícias) recuperadas pelo sistema de recuperação de informação

CORREIO BRAZILIENSE Busca

Pesquisa avançada ▲

Filtrar por Data:

 até 

Filtrar por seções:

- | | |
|--|--|
| <input type="checkbox"/> Política Brasil | <input type="checkbox"/> Mundo |
| <input type="checkbox"/> Economia | <input type="checkbox"/> Diversão e Arte |
| <input type="checkbox"/> Especiais | <input type="checkbox"/> Revista |
| <input type="checkbox"/> Capa do Dia | <input type="checkbox"/> Carnaval2016 |
| <input type="checkbox"/> Política | <input type="checkbox"/> CorreioDebate |
| <input type="checkbox"/> Brasil | <input type="checkbox"/> Canta Brasília |
| <input type="checkbox"/> Meu Lugar | <input type="checkbox"/> cbbrands |
| <input type="checkbox"/> Comunidade Ei | <input type="checkbox"/> VidaEscolar |
| <input type="checkbox"/> Summit | <input type="checkbox"/> Bbb18 |

Ordenar por:

- ▲
-
-
- Ciência e Saúde
- Turismo
- Carnaval2018
- CNC
- Parceiros
- diadospais
- Correio Talks
- Ultimas noticias

Autor:

-
- Economia
 - Tecnologia
 - Eu Estudante
 - Carnaval
 - ForumMundialdaAgua
 - Escolhaaescola
 - Carnaval2017
 - Dialogar para liderar
 - Eleições

Fig. 4. Filtros adicionais e opcionais para refinamento da recuperação da informação