



TENDÊNCIAS ATUAIS E PERSPETIVAS FUTURAS EM ORGANIZAÇÃO DO CONHECIMENTO

ATAS DO III CONGRESSO ISKO ESPANHA-PORTUGAL
XIII CONGRESSO ISKO ESPANHA

Universidade de Coimbra, 23 e 24 de novembro de 2017

Com a coordenação de

Maria da Graça Simões, Maria Manuel Borges

TÍTULO

Tendências Atuais e Perspetivas Futuras em Organização do Conhecimento: atas do III Congresso ISKO Espanha e Portugal - XIII Congresso ISKO Espanha

COORDENADORES

Maria da Graça Simões
Maria Manuel Borges

EDIÇÃO

Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20

ISBN

978-972-8627-75-1

ACESSO

<https://purl.org/sci/atas/isko2017>

COPYRIGHT

Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

OBRA PUBLICADA COM O APOIO DE



FLUC FACULDADE DE LETRAS
UNIVERSIDADE DE COIMBRA



CEIS 20
CENTRO DE ESTUDOS
INTERDISCIPLINARES
DO SÉCULO XX
UNIVERSIDADE DE COIMBRA

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

PROJETO UID/HIS/00460/2013



AVALIAÇÃO DOS MODELOS CLÁSSICOS DE RECUPERAÇÃO DA INFORMAÇÃO EMPREGADOS NA ÀREA DA SAÚDE

Josiane Mello¹, Angel Freddy Godoy Viera²

¹Doutoranda do Programa de Pós-Graduação em Ciência da Informação – UFSC, mellojosi@hotmail.com

²Professor Doutor do Programa de Pós-Graduação em Ciência da Informação – UFSC, Líder do Grupo de Pesquisa RITA – Recuperação da Informação e Tecnologias Avançadas, a.godoy@ufsc.br

RESUMO O presente artigo tem por objetivo avaliar o emprego dos modelos clássicos de recuperação da informação na área da saúde, por meio da publicação científica. Para alcance de tal objetivo, realizou-se um levantamento bibliográfico, no portal de periódicos da CAPES, em bases de dados da área da saúde. Trata-se de uma pesquisa exploratória-descritiva e bibliográfica, elaborada a partir de material já publicado e disponibilizado nas bases de pesquisa. Os resultados mostram que os modelos clássicos de recuperação da informação, se inserem de forma positiva na área estudada, contribuindo diretamente na vida da população, seja por meio de diagnósticos rápidos e precisos ou por métodos e técnicas inovadoras de prevenção e tratamento de enfermidades e etc. O modelo clássico Booleano pode ser empregado na Fonoaudiologia para classificar a gravidade de distúrbios articulatorios, na Oncologia e entre outras especialidades da área da saúde para acompanhar o tratamento de enfermidades, pela simplicidade e por se tratar de um modelo intuitivo com semântica precisa. Já o modelo clássico vetorial pode auxiliar no diagnóstico, prevenção, controle e tratamento de enfermidades. Por fim, o modelo probabilístico pode ser empregado na área da saúde, promovendo interação entre uma máquina e um sistema de informação de uma unidade médica e para auxiliar no processo de tomada de decisão médica. Recomenda-se que essa pesquisa seja ampliada, incluindo os modelos alternativos de RI e as tecnologias de máquinas de busca.

PALAVRAS-CHAVE *Recuperação da Informação, Sistemas de Recuperação da Informação, Modelos de Recuperação da Informação, Saúde.*

ABSTRACT The objective of this paper is to evaluate the use of classic models of information retrieval in the health area, through scientific publication. In order to reach this objective, a bibliographic survey was carried out, in the CAPES journals portal, in databases of the health area. It is an exploratory-descriptive and bibliographical research, elaborated from material already published and made available in the research bases. The results show that the classic models of information retrieval are inserted positively in the studied area, contributing directly to the population's life, either through rapid and precise diagnoses or innovative methods and techniques of prevention and treatment of diseases and etc. The classic Boolean model can be used in Speech Therapy to classify the severity of articulatory disorders, in Oncology and among other specialties of the health area to accompany the treatment of diseases, for simplicity and for being an intuitive model with precise semantics. The classic vector model can help in the diagnosis, prevention, control and treatment of diseases. Finally, the probabilistic model can be used in the health area, promoting interaction between a machine and an information system of a medical unit and to assist in the medical decision-making process. It is recommended that this research be expanded, including alternative models of RI and search engine technologies.

KEYWORDS *Information Retrieval, Information Retrieval Systems, Information Retrieval Models, Health.*

COPYRIGHT Este trabalho está licenciado com uma Licença Creative Commons - Atribuição 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/deed.pt>)

INTRODUÇÃO

A explosão informacional ocasionou vários problemas relacionados ao gerenciamento da informação, conforme descrito por Bush, em 1945, no artigo intitulado “As We May Think”. Segundo este autor, uma possível solução para tais problemas seria a automatização dos processos de recuperação da informação. Para tanto, ele chegou a imaginar um mecanismo específico, denominado de Memex (Memory Extension), que agregava as mais modernas tecnologias de informação existentes na época.

Na década seguinte, Mooers propôs a criação da disciplina de Recuperação da Informação, visando resolver os problemas oriundos do crescimento desenfreado do número de documentos produzidos e o acesso a estes.

Hoje, ainda enfrentam-se tais problemas e em uma escala muito maior, e, como Bush, busca-se na tecnologia a solução para tais problemas (FERNEDA, 2012).

Nos últimos anos, a área de Recuperação da Informação se desenvolveu significativamente, seus objetivos iniciais de indexação de textos e de busca por documentos úteis foram ampliados, incorporando, na atualidade, a pesquisa em modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuário, filtragem, processamento de linguagem natural e entre outros.

Em se tratando de pesquisa, de acordo com Baeza-Yates e Ribeiro-Neto (2013) e Xie (2008), a referida área está contemplando estudos com dois enfoques distintos e complementares: um centrado no computador e outro centrado no usuário. Nos estudos com o enfoque centrado no computador, a recuperação da informação versa principalmente na construção de índices eficientes, no processamento de consultas com alto desempenho e no desenvolvimento de algoritmos de ranqueamento, objetivando melhorar os resultados. No que tange aos estudos com enfoque centrado no usuário, a recuperação da informação versa principalmente em investigar o comportamento do usuário, entender suas principais necessidades e determinar como esse entendimento afeta a organização e a operação do sistema de recuperação da informação.

No caso desta pesquisa, trata-se de um estudo centrado no computador, com o objetivo de investigar como os Modelos Clássicos de Recuperação da Informação são empregados na área da saúde e quais as contribuições para a área supracitada?

A escolha pela área da saúde se deu a partir de dois motivos. O primeiro em função de a referida área ter se desenvolvido substancialmente, sobretudo, depois do advento das tecnologias da informação e comunicação e as utilizaram amplamente. E o segundo, a partir do entendimento do papel e urgência da informação neste espaço. Sabe-se que, nesta área, a necessidade de informação para a tomada de decisão se faz presente constantemente, e a exatidão, pertinência, precisão e o tempo que esta leva para chegar às mãos do tomador de decisão fazem toda a diferença, podendo ter implicações no controle e no agravo de enfermidades, reduzir o número de óbitos, trazer métodos e técnicas inovadoras para realizações de exames, diagnósticos e tratamentos, além de possibilitar a construção de novos conhecimentos aos profissionais desta área.

RECUPERAÇÃO DA INFORMAÇÃO

A Recuperação da Informação, doravante RI é uma área vasta dentro da Ciência da Computação e Ciência da Informação, pode-se entender a RI como um processo no qual um usuário converte sua necessidade de informação numa expressão de busca que submete ao sistema e o mesmo responde com uma listagem de registros que são relevantes para essa necessidade de informação.

Conforme relatos de Baeza-Yates e Ribeiro-Neto (2013), a preocupação em organizar para posteriormente buscar e recuperar a informação não é algo recente, as primeiras bibliotecas foram criadas para atender este objetivo, entre 3000 e 2000 a.C, sendo também, as primeiras instituições a adotarem sistemas de RI para recuperar informações.

Sobre isso, Foresti e Godoy Viera (2016, p. 6), pontuam que,

tradicionalmente a Recuperação da Informação (RI) acontecia nos ambientes escolares e acadêmicos em bibliotecas, local em que o usuário atuava na sua ânsia por informação e conhecimento novo manipulando as vetustas fichas catalográficas em formato papel. Suas necessidades eram grandemente orientadas pelo que acontecia em sala de aula. Mas não eram apenas as bibliotecas que permitiam a busca de informação em suas coleções, os arquivos e museus há muito atendem pesquisadores, empresas e a sociedade em geral através de seus inventários de demais instrumentos de pesquisa.

Com o passar do tempo, para acompanhar as transformações paradigmáticas da sociedade, a RI foi ampliando sua abrangência, com a banalização dos computadores e o advento das tecnologias móveis, a RI em dispositivos móveis cresceu significativamente, conforme pesquisa recente do IBGE, abordada no estudo de Foresti e Godoy Viera (2016).

Concomitante com a ampliação da abrangência, o conceito da RI sofreu alterações, deixando de ser vista como sinônimo de busca de literatura, conforme definição de Lancaster (1978), incorporando aspectos lógicos, Linguísticos e tecnológicos em sua conceituação.

Na atualidade, a RI é definida, conforme Baeza-Yates; Ribeiro-Neto (2013, p.1), como uma área, que,

trata de representação, armazenamento, organização e acesso a itens de informação, como documentos, páginas *web*, catálogos online, registros estruturados e semiestruturados, objetos multimídia, etc. A representação e a organização dos itens de informação devem fornecer aos usuários facilidade de acesso às informações de seu interesse (BAEZA-YATES; RIBEIRO-NETO, 2013, p.1).

Devido ao grande volume de documentos não estruturados armazenados na internet, surgem vários problemas relacionados à recuperação da informação. Assim, de acordo com Sharma (2013), a referida disciplina necessita de métodos que recuperem a informação de maneira eficaz, de forma automatizada.

Para atingir esses objetivos, os sistemas de RI fazem uso de modelos de recuperação da informação, que são apresentados na seguinte seção.

MODELOS DE RECUPERAÇÃO DA INFORMAÇÃO

Com as demandas e desafios impostos pela sociedade hodierna, os usuários necessitam cada vez mais de informações de alto nível de complexidade, a fim de empregá-las em suas atividades laborais ou em benefício próprio, suprimindo uma necessidade informacional imediata.

Essa necessidade de informação é traduzida pelo usuário em uma consulta, ou em uma sequência de consultas, que é submetida ao sistema de RI, sob um elenco de palavras chave, ou termos indexados, visando a obtenção de uma lista de documentos úteis e relevantes para ele.

Na percepção de Baeza-Yates e Ribeiro-Neto (2013) um sistema de RI deve ser capaz de recuperar todos os documentos que são relevantes à necessidade informacional do usuário e, ao mesmo, recuperar o menor número possível de documentos irrelevantes.

No entanto, é utopia esperar que um sistema de RI atenda plenamente tal objetivo, pois conforme observado pelos autores supracitados, a relevância está ligada a um julgamento pessoal, que varia de acordo com a tarefa a ser realizada e com o contexto desta. A relevância pode mudar com o tempo, à medida que novas informações estão disponíveis; com o local, talvez a resposta mais relevante seja a mais próxima e ainda com o dispositivo, considerando que a melhor resposta pode ser aquela disponibilizada em um documento pequeno, sendo baixada e visualizada mais rápido.

Diante da impossibilidade de um sistema de RI fornecer apenas respostas relevantes a todos os usuários o tempo todo, foram incorporados a estes, os modelos de RI, visando aproximar a necessidade de informação do usuário com o resultado obtido na busca, por meio do emprego de algoritmos e premissas de relevância definida em cada modelo de RI.

Os primeiros modelos de RI propostos foram os clássicos: Booleano, Vetorial e Probabilístico. Nestes modelos, “cada documento é descrito por um conjunto de palavras-chave representativas – também chamadas de termos de indexação – que busca representar o assunto do documento e sumarizar seu conteúdo de forma significativa”. (SOUZA, 2006, p.166).

De forma sucinta, apresentam-se a funcionalidade e desempenho dos modelos supracitados: O Modelo booleano é baseado na teoria de conjuntos e na álgebra Booleana, conforme descrição de Silva, Santos e Ferneda (2013, p.29-30),

o conteúdo informacional dos documentos é representado por um conjunto de termos de indexação. As buscas são formuladas por meio de uma expressão booleana composta por termos ligados através dos operadores lógicos (AND, OR e NOT). O resultado de uma busca é composto por um conjunto de documentos cuja representação satisfaz às restrições lógicas da expressão de busca. (SILVA, SANTOS E FERNEDA, 2013, p. 29 - 30).

Baeza-Yates e Ribeiro-Neto (2013, p.31) ressaltam como vantagens e desvantagens do modelo Booleano:

[...] o formalismo claro por trás do modelo e sua simplicidade, com a adoção de pesos binários para os termos de indexação. As principais desvantagens são o fato de não existir um ranqueamento, o que pode levar à recuperação de muitos ou de poucos documentos. Além disso, a formulação de consultas Booleanas é inconveniente para a maior parte dos usuários. Atualmente, sabe-se que a ponderação dos termos de indexação pode levar a uma melhoria substancial na qualidade da recuperação [...]. (BAEZA-YATES e RIBEIRO-NETO, 2013, p.31).

O Modelo vetorial de acordo com Baeza-Yates e Ribeiro-Neto (2013) considera casamentos parciais do termo de consulta, por meio de atribuição de pesos não binários aos termos de indexação das consultas e dos documentos, no qual, são usados para computar o grau de similaridade entre cada documento armazenado no sistema e a consulta do usuário.

A lista de resultados dos documentos recuperados é exibida em ordem decrescente, de acordo com o grau de similaridade destes, de modo a fornecer uma resposta mais precisa ao usuário, quando comparado ao modelo Booleano, em função do ranqueamento de documentos.

O modelo vetorial apresenta as seguintes vantagens, conforme Baeza-Yates e Ribeiro-Neto (2013, p.47),

- Seu esquema de ponderação de termos melhora a qualidade da recuperação da informação;
- Sua estratégia de casamento parcial permite a recuperação de documentos que aproximam as condições da consulta;
- A fórmula do cosseno ordena os documentos de acordo com o seu grau de similaridade em relação à consulta;
- A normalização pelo tamanho do documento está naturalmente embutida no modelo.

Como desvantagem o modelo não considera as dependências entre os termos de indexação. Apesar disso, o modelo não deixa de ser apontado como um excelente recurso para o ranqueamento de coleções genéricas.

Já o modelo probabilístico, desenvolvido por Robertson e Sparck Jones em 1976, faz uso da teoria probabilística para resolver o problema da RI. Souza (2006, p.167) diz que,

nesse modelo, supõe-se que exista um conjunto ideal de documentos que satisfaz a cada uma das consultas ao sistema, e que este conjunto pode ser recuperado. Através de tentativa inicial com um conjunto de documentos (para a qual se podem utilizar técnicas de outros modelos, como o vetorial) e do feedback do usuário em sucessivas interações, buscas e aproximar cada vez mais deste conjunto ideal, por meio de análise dos documentos considerados pertinentes pelo usuário. O valor desse modelo está em considerar a interação contínua com o usuário como um caminho para refinar o resultado continuamente.

No que tange as vantagens e desvantagens deste modelo, Baeza-Yates e Ribeiro-Neto (2013) pontuam algumas, respectivamente, que teoricamente, trata-se de um ótimo modelo de RI, por ranquear os documentos de forma decrescente, pautado na probabilidade de serem relevantes, baseado na informação disponível ao sistema. Embora, na prática, isso não funciona tão bem, pois a relevância do documento é afetada por variáveis externas ao sistema. Como inconveniente, o modelo necessita que seja estimado a separação inicial dos documentos em conjuntos de relevantes e não relevante. O tamanho dos documentos não é normalizado e o método também não leva em consideração a frequência na qual um termo de indexação ocorre em um documento.

METODOLOGIA

Esta pesquisa caracteriza-se quanto aos objetivos como exploratória-descritiva e quanto aos procedimentos técnicos como bibliográfica.

De acordo com Gil (2008, p. 27), as pesquisas exploratórias:

[...] são desenvolvidas com o objetivo de proporcionar visão geral, do tipo aproximativo, acerca de determinado fato. Este tipo de pesquisa é realizado especialmente quando o tema escolhido é pouco explorado e torna-se difícil sobre ele formular hipóteses precisas e operacionalizáveis.

Já as pesquisas descritivas “[...] têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis.” (GIL, 2008, p. 28).

Quanto aos procedimentos técnicos, define-se como uma pesquisa bibliográfica, realizada nas bases de dados da área da saúde: Biblioteca Virtual da Saúde (BVS) - BIREME, MEDLINE, LILACS, IBECs do portal de periódicos da CAPES (<http://www.capes.gov.br/>), com o objetivo de levantar estudos, que empregam os modelos clássicos de recuperação da informação no domínio da saúde.

Para tal, a cobertura temporal utilizada foi de cinco anos (2012-2016), tipo de documento: artigo, idiomas: inglês, espanhol e português, campo de busca: título, resumo e assunto; assunto principal: armazenamento e recuperação da informação, limite da pesquisa: humanos e texto completo disponível.

Realizaram-se busca avançada, com auxílio de operadores booleanos. A estratégia de busca empregada é exibida no quadro 1.

Quadro 1 – estratégias de busca utilizadas

Estratégia de busca	Operador booleano	Total de documentos recuperados	Documentos relevantes
“information retrieval” AND model AND (Boolean OR Probabilistic OR vector) AND (medical OR medicine OR health)	AND/OR	12	07

Fonte: Mello e Godoy Viera (2017).

A pesquisa recuperou 12 documentos, destes, fez-se a leitura e excluíram-se 05 estudos, pois não estavam alinhados com o escopo desta pesquisa. Destes, 03 abordavam a técnica de aprendizado de máquina, 01 abordava o modelo alternativo de RI – Redes Bayesianas e 01 abordava o processo de elaboração de ontologia para auxiliar na atividade de processamento da linguagem natural. Os 07 documentos relevantes, que compõem o corpus desta pesquisa, encontram-se referenciados ao final deste resumo alargado, intitulado como referência do corpus da pesquisa.

RESULTADOS

Constatou-se por meio da literatura especializada o emprego de todos os modelos clássicos de recuperação da informação na área da saúde, os estudos foram enumerados de 01 até 07, conforme exibição no quadro 2

O modelo clássico Booleano foi identificado nos estudos 2 e 4. No primeiro mencionado, para

classificar a gravidade do desvio fonológico (DF). Já o segundo abordava o emprego do modelo na construção de uma rede para acompanhamento do tratamento de câncer de pulmão (SCLC). Verifica-

se com base nestas aplicações, que o modelo clássico Booleano pode ser empregado na Fonoaudiologia para classificar a gravidade de distúrbios articulatorios, na Oncologia e entre outras especialidades da área da saúde para acompanhar o tratamento de enfermidades, pela simplicidade e por se tratar de um modelo intuitivo com semântica precisa.

Já o modelo clássico vetorial foi sinalizado nos estudos 03 e 07. Onde o primeiro mencionado identificou o emprego do modelo supramencionado com foco na recuperação de caso de atendimento similar, para auxiliar no cuidado clínico e na pesquisa. Enquanto o segundo estudo foi aplicado para descobrir relações e outros significados entre drogas e doenças, reconhecimento de entidade multi-nível (gene, fármaco, doença, sintoma) e extração de interação (medicamento-gene, doença genética) da literatura. Verifica-se neste caso, uma importante contribuição do modelo supramencionado na área da saúde, podendo este, auxiliar no diagnóstico, prevenção, controle e tratamento de enfermidades.

Por fim, o modelo probabilístico foi identificado nos estudos 01, 05, 06. O primeiro estudo detectou o emprego do modelo em questão para o desenvolvimento de um componente de comunicação entre máquinas de anestesia e o sistema de informação de um determinado hospital. O segundo estudo supramencionado foi empregado para calcular e interpretar o atraso mútuo da informação (TDMI) para uma tarefa complexa, diversa e população escassamente medida, possivelmente não-estacionárias de séries temporais de composição de origem desconhecida. O terceiro estudo apresenta o emprego do modelo em tela para precisar se um documento biomédico acessado no passado poderá ser de interesse de acesso no futuro. De modo que, essa métrica de desejabilidade fornece um meio para melhorar a classificação de documentos para sistemas recuperação da informação e uma base para o avanço de modelos de recuperação da informação Bayesianos. Verifica-se que esse modelo pode ser empregado na área da saúde, promovendo interação entre uma máquina e um sistema de informação de uma unidade médica e para auxiliar no processo de tomada de decisão médica.

Quadro 2 – Resumo da aplicação dos modelos e tecnologias de RI

Número do Estudo	Modelos clássicos de RI
01 González Martínez, Y. ; Castillo García, M. ; Quevedo, Y. S. (2014)	Fez uso do modelo probabilístico com o apoio de agentes inteligentes, uso de padrões de vídeos e codificação de vídeos, metadados e técnicas de processamento de conteúdo para o desenvolvimento de um componente de comunicação entre máquinas de anestesia e o sistema de informação de um determinado hospital.
02 Brancalioni, A. R. ; Magnago, K. F. ; Keske-soares, M. (2012)	Empregou o modelo de recuperação da informação booleano , utilizando a teoria de conjunto Fuzzy para classificar a gravidade do desvio fonológico (DF).
03 Moen H; Ginter F; Marsi E; Peltonen L-M; Salakoski T; Salanterä S. (2015)	Relata o emprego de vários métodos para a recuperação de informação, com foco na recuperação de caso de atendimento, com base na similaridade textual, onde a similaridade é medida através da modelagem específica do domínio da semântica distributiva das palavras. Dentre os modelos empregados, identificou-se o modelo vetorial , com variantes de indexação aleatória e o modelo de rede neural semântica word2vec.
04 Udyavar, A. R; Wooten, D. J; Hoeksema, M. D; Bansal, M.; Califano, A.; Estrada, L; Chnell, S. ; Irish, J.	O câncer de pulmão (SCLC) é uma doença devastadora devido à sua propensão para a invasão precoce e recaída refratária após a resposta ao tratamento inicial. A topologia desta rede de SCLC fatores de transcrição (FT) foi derivada de conhecimento prévio e simulada usando modelagem booleana . Estas simulações predisseram que a rede se instala em atratores, ou padrões de expressão de (FT), que se correlacionam com os fenótipos (NE) ou (ML), sugerindo que a dinâmica da rede (FT) subjaz ao surgimento de fenótipos heterogêneos de SCLC.

<p>M; Massion, P. P; Quaranta, V. (2016)</p>	
<p>05 Albersa, D. J. ; Hripcsakb, G. (2012)</p>	<p>O estudo apresenta o emprego do modelo probabilístico, para calcular e interpretar o atraso mútuo da informação (TDMI) para uma tarefa complexa, diversa e população escassamente medida, possivelmente não-estacionárias de séries temporais de composição desconhecida e de origem. O principal veículo utilizado para esta análise é uma comparação entre a informação mútua temporizada, em média, a população e o atraso de informação mútua de uma população de agregados (aqui, agregação implica que a população é em conjunto antes de quaisquer estimativas estatísticas são implementadas). Através do uso de ferramentas de informação teórica, uma sequência de cálculos praticamente exequível são detalhados que permitem para a média e a agregação temporizada informação mútua ser interpretado. Além disso, esses cálculos também podem ser usados para compreender o grau de homo ou heterogeneidade presente na população. Os métodos foram aplicados e demonstrados na série de tempo das medições de glicose de duas subpopulações diferentes dos indivíduos do repositório de registro eletrônico de saúde da University Medical Center of Columbia, revelando uma imagem da composição da população, bem como características fisiológicas.</p>
<p>06 Goodwin, J. C. ; Johnson, T. R.; Cohen, T. ; Herskovic, J. R; Bernstam, E. V. (2012)</p>	<p>O artigo apresentou o emprego do modelo probabilístico para precisar se o documento biomédico acessado no passado poderá ser de interesse de acesso no futuro. Para tal, os autores avaliaram dois modelos de acesso ao documento, baseados em Anderson e Schooler (1991) o primeiro modelo baseado na frequência com que um documento foi acessado. O segundo baseado na frequência e relevância . O estudo conclui que é possível modelar com precisão a probabilidade de um documento ser acessado no futuro com base em acessos anteriores. O emprego deste modelo torna-se relevante diante da sobrecarga de informação no domínio biomédico, o esforço significativo de pesquisa centrou-se nos sistemas de recuperação de informação (RI). A métrica de desajustabilidade fornece um meio para melhorar a classificação de documentos para sistemas IR e uma base para o avanço de modelos bayesianos RI, fornecendo uma estimativa de probabilidade anterior teoricamente motivada.</p>
<p>07 Lee, S. ; Choi, J. ; Park, K. ; Song, M.; Lee, D. (2012)</p>	<p>O estudo utiliza o modelo vetorial, denominado de Swanson ABC para inferir relações ocultas na literatura biológica. Trata-se de um modelo consistente, no entanto, o modelo é inadequado para inferir relações com informações de contexto. Além disso, o modelo gera uma grande quantidade de candidatos a partir de texto biológico, e é uma técnica semi-automática, de mão-de-obra intensiva, que exige a entrada manual do especialista humano. Para enfrentar esses problemas, foram incorporado termos de contexto para inferir as relações entre as interações AB e as interações do BC. Foram propostas 3 etapas para descobrir relacionamentos ocultos significativos entre drogas e doenças: 1) reconhecimento de entidade multi-nível (gene, fármaco, doença, sintoma), 2) extração de interação (medicamento-gene, doença genética) da literatura, 3) contexto cálculo de pontuação de similaridade baseada em vetores. Posteriormente, foi avaliado a hipótese com os conjuntos de dados da "Doença de Alzheimer" relacionados 77.711 resumos da PubMed. Como padrões de ouro, os bancos de dados PharmGKB e CTD foram usados. Os resultados indicam que a inferência de relação baseada em contexto alcançou uma melhor precisão do que a abordagem do modelo ABC anterior. A análise da literatura também mostra que as interações inferidas pela abordagem baseada no contexto são mais significativas do que as interações pelo modelo ABC anterior.</p>

Fonte: Mello e Godoy Viera (2017)

CONSIDERAÇÕES FINAIS

A saúde é uma área que tem se desenvolvido substancialmente, sobretudo, com o auxílio das tecnologias da informação e comunicação.

Identificou-se o emprego de todos os modelos clássicos de recuperação da informação no domínio da saúde.

Pela pesquisa desenvolvida foi possível constatar que tais modelos, quando aplicados na área supracitada, proporcionam vários benefícios para a vida da população: diagnósticos rápidos e precisos, métodos e técnicas inovadoras de prevenção e tratamento de enfermidades e etc.

O modelo clássico Booleano pode ser empregado na Fonoaudiologia para classificar a gravidade de distúrbios articulatorios, na Oncologia e entre outras especificidades da área da saúde para

acompanhar o tratamento de enfermidades, pela simplicidade e por se tratar de um modelo intuitivo com semântica precisa.

Já o modelo clássico vetorial pode auxiliar no diagnóstico, prevenção, controle e tratamento de enfermidades.

Por fim, o modelo probabilístico pode ser empregado na área da saúde, promovendo interação entre uma máquina e um sistema de informação de uma unidade médica e para auxiliar no processo de tomada de decisão médica.

Recomenda-se ampliar o escopo da pesquisa, incluindo os modelos alternativos de recuperação da informação e as tecnologias de máquina de busca no domínio da saúde.

REFERÊNCIAS BIBLIOGRÁFICAS DO CORPUS DA PESQUISA

ESTUDO 1

González Martínez, Y. ; Castillo García, M. ; Quevedo, Y. S. (2014). Componente para la lectura de datos por alas-HIS desde máquinas de anestesia. *RCIM [online]*, 6 (1), n.1, 110 - 117.

ESTUDO 2

Brancalioni, A. R. ; Magnago, K. F. ; Keske-soares, M. (2012). Validation of a fuzzy linguistic model to classify the severity of phonological disorder, *Revista CEFAC: Atualização Científica em Fonoaudiologia e Educação*, 14 (3), 448-458.

ESTUDO 3

Moen H; Ginter F; Marsi E; Peltonen L-M; Salakoski T; Salanterä S. (2015). Care episode retrieval: distributional semantic models for information retrieval in the clinical domain. *BMC Medical Informatics and Decision Making*, 15 (2), 1-19.

ESTUDO 4

Udyavar, A. R; Wooten, D. J; Hoeksema, M. D; Bansal, M.; Califano, A.; Estrada, L; Chnell, S. ; Irish, J. M; Massion, P. P; Quaranta, V. (2016). Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. *Cancer Res.* 77(5), 1063-1074.

ESTUDO 5

Albersa, D. J. ; Hripcsakb, G. (2012). Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations. *AIP Chaos:an Interdisciplinare Journal of Nonlinear Science*, 22(1), 1-26.

ESTUDO 6

Goodwin, J. C. ; Johnson, T. R.; Cohen, T. ; Herskovic, J. R; Bernstam, E. V. (2012). Predicting biomedical document access as a function of past use. *J Am Med Inform Assoc*; 19(3), 473- 478.

ESTUDO 7

Lee, S. ; Choi, J. ; Park, K. ; Song, M.; Lee, D. (2012). Discovering context-specific relationships from biological literature by using multi-level context terms. *BMC Medical Informatics And Decision Making*, 12(1), 1-12.

REFERÊNCIAS BIBLIOGRÁFICAS

Baeza-Yates, R.; Ribeiro-Neto, B. (2013). *Recuperação de informação: conceitos e tecnologia das máquinas de busca*. 2. ed. Porto Alegre: Bookman.

Ferneda, Edberto. (2012). *A. Introdução aos modelos computacionais de recuperação de informação*. Rio de Janeiro: Ciência Moderna.

Foresti, F. ; Godoy Vieira, A. F. (2016). A Recuperação da informação em dispositivos móveis, *Biblionline*, João Pessoa, 12 (2), 6 – 25.

Gil, Antônio Carlos. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2008.

Lancaster, F. W. (1978). *Information retrieval systems: characteristics, testing and evaluation*. 2. ed. New York, NY: Wiley.

Sharma, A. (2013). Intelligent information retrieval system: a survey. *Advance in Electronic and Electric Engineering*. 3 (1).

Silva, R. E. ; Santos, P. L. V. A. C. ; FERNEDA, E. (2013). Modelos de recuperação de informação e web semântica: a questão da relevância. *Inf. Inf.*, Londrina, 18 (3), 27 – 44.

Souza, R. R. (2016). Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, Belo Horizonte, 11 (2), 161-173.

Xie, I. (2008). *Interactive information retrieval in digital environments*. New York: IGI Publishing.