

Mapeando a dinâmica da informática médica: uma análise bibliométrica do campo científico

Mapping the dynamics of medical informatics: a bibliometric analysis of the scientific field

Mapeo de la dinámica de la informática médica: un análisis bibliométrico del campo científico

Bruno Elias Penteadó^{1,a}

bpentead@gmail.com | <https://orcid.org/0000-0002-8366-5512>

Marcelo Fornazin^{2,3,b}

marcelo.fornazin@ensp.fiocruz.br | <https://orcid.org/0000-0002-0379-5801>

Leonardo Castro^{2,c}

leonardo.castro@ensp.fiocruz.br | <http://orcid.org/0000-0002-9376-5103>

Sandro Luis Freire^{4,d}

sandrofreire@gmail.com | <https://orcid.org/0000-0002-4532-5218>

¹ Fundação Oswaldo Cruz, Estratégia Fiocruz para a Agenda 2030. Rio de Janeiro, RJ, Brasil.

² Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca. Rio de Janeiro, RJ, Brasil.

³ Universidade Federal Fluminense. Niterói, RJ, Brasil.

⁴ Instituto Nacional de Câncer. Rio de Janeiro, RJ, Brasil.

^a Doutorado em Ciência da Computação pela Universidade de São Paulo.

^b Doutorado em Administração pela Fundação Getúlio Vargas.

^c Doutorado em Antropologia pela Universidade Federal do Rio de Janeiro.

^d Mestrado em Gestão e Estratégia pela Universidade Federal Rural do Rio de Janeiro.

RESUMO

A saúde digital é um assunto emergente em fóruns acadêmicos, nas políticas públicas e nas organizações de saúde. Supondo que a saúde digital deriva de conhecimentos da informática médica, este artigo apresenta resultados de uma pesquisa bibliométrica sobre a evolução conceitual e tecnológica do campo da informática médica nas últimas décadas, enfatizando aspectos metodológicos. O trabalho realizou bibliometria em metadados de 100 mil artigos indexados sob a categoria ‘*medical informatics*’ na base de dados Web of Science entre os anos de 1960 e 2020. Foram realizadas análises longitudinais com utilização dos *softwares* Bibliometrix e CorText em três eixos: quantidade de publicações, países dos autores e palavras-chave. Conforme a hipótese metodológica que orientou o estudo, as mudanças terminológicas verificadas ao longo do tempo oferecem uma visão aproximativa das mudanças conceituais e tecnológicas do campo de pesquisa da informática médica. Os resultados mostram que esse campo de investigação apresentou crescimento consistente ao longo das últimas seis décadas, expandindo-se para diferentes países. As mudanças terminológicas e conceituais detectadas pela análise de palavras-chave permitiram a identificação de períodos temporais definidos, associados a rótulos genéricos como ‘*health informatics*’, ‘*e-health*’. O rótulo ‘*medical informatics*’ é recorrente como termo mais geral a designar o campo de aplicação, em razão de sua adoção por associações científicas internacionais a partir da década de 1970. Nos últimos cinco anos, pode-se identificar a emergência do termo ‘*digital health*’, que possivelmente será o conceito dominante na década que se inicia. A análise de palavras-chave também indica a associação entre mudanças terminológicas e de tecnologias, o que reforça as relações entre conceitos e aplicações tecnológicas de cada período.

Palavras-chave: Bibliometria; Informática médica; Informática em saúde; Saúde eletrônica; Saúde digital.

ABSTRACT

Digital health is an emerging topic in academic forums, public policies, and healthcare organizations. Assuming that digital health derives from previous medical informatics knowledge, this work presents findings from a bibliometric study on medical informatics technological and conceptual evolution in the last decades, emphasizing methodological aspects. We performed a bibliometric analysis in metadata from 100,000 papers indexed under the category 'medical informatics' in the Web of Science database between 1960 and 2020. Longitudinal analysis using software Bibliometrix and CorText were conducted in three axes: frequency of items, authors' countries, and keywords. Based on the methodological hypothesis guiding the study, the changes in keywords over time offer a proxy view on the conceptual and technological changes in the medical informatics research field. The results show that medical informatics consistently grew over the last six decades, expanding to several countries. Conceptual and technological changes that emerged from the keyword analysis supported the identification of well delimited periods related to general labels, such as: 'health informatics' and 'e-health'. The 'medical informatics' is recurring as a general label due to international scientific associations' adoption since the early 1970s. Moreover, in the last five years, we could identify the term 'digital health', which will probably be a major label in the next decade. The keyword analysis also showed the association between labels and technological changes, adding more evidence that these changes are related to concepts and technological applications.

Keywords: Bibliometrics; Medical informatics; Health informatics; Electronic health; Digital health.

RESUMEN

La salud digital es un tema emergente en foros académicos, políticas públicas y organizaciones de salud. Asumiendo que la salud digital deriva del conocimiento de la informática médica, este artículo presenta los resultados de una investigación bibliométrica sobre la evolución conceptual y tecnológica del campo de la informática médica en las últimas décadas, enfatizando aspectos metodológicos. El trabajo realizó bibliometría en metadatos de 100.000 artículos indexados en la categoría 'informática médica' en la base de datos Web of Science entre los años 1960 y 2020. Se realizaron análisis longitudinales utilizando los *softwares* Bibliometrix y CorText en tres ejes: número de publicaciones, nacionalidad de los autores y palabras clave. De acuerdo con la hipótesis metodológica que orientó el estudio, los cambios terminológicos verificados a lo largo del tiempo ofrecen una visión aproximada de los cambios conceptuales y tecnológicos en el campo de la investigación en informática médica. Los resultados muestran que este campo de investigación ha mostrado un crecimiento constante durante las últimas seis décadas, expandiéndose a diferentes países. Los cambios terminológicos y conceptuales detectados por el análisis de palabras clave permitieron identificar periodos de tiempo definidos, asociados a etiquetas genéricas como 'informática en salud', 'e-salud'. La etiqueta 'informática médica' es recurrente como el término más general para designar el campo de aplicación, debido a su adopción por parte de asociaciones científicas internacionales desde la década de 1970. En los últimos cinco años se puede identificar el surgimiento del término 'salud digital', que posiblemente será el concepto dominante en la próxima década. El análisis de palabras clave también indica la asociación entre cambios terminológicos y tecnológicos, lo que refuerza las relaciones entre conceptos y aplicaciones tecnológicas en cada período.

Palabras clave: Bibliometría; Informática médica; Informática de la salud; Salud electrónica; Salud digital.

Este artigo compõe o dossiê Estudos métricos da informação científica em saúde, parte 1.

Contribuição dos autores:

Concepção e desenho do estudo: Bruno Elias Penteado, Marcelo Fornazin, Leonardo Castro, Sandro Luis Freire.

Aquisição, análise ou interpretação dos dados: Bruno Elias Penteado, Marcelo Fornazin, Leonardo Castro.

Redação do manuscrito: Bruno Elias Penteado, Marcelo Fornazin, Leonardo Castro.

Revisão crítica do conteúdo intelectual: Bruno Elias Penteado, Marcelo Fornazin, Leonardo Castro, Sandro Luis Freire.

Declaração de conflito de interesses: não há.

Fontes de financiamento: Este trabalho foi financiado pela Estratégia Fiocruz para a Agenda 2030.

Considerações éticas: não há.

Agradecimentos/Contribuições adicionais: Agradecemos à Estratégia Fiocruz para a Agenda 2030 pelo apoio à realização desta pesquisa. Agradecemos a Fabio Gouveia e Alexandre Hannud Abdo pelos comentários nas etapas preliminares da elaboração deste trabalho. Os autores gostariam de agradecer também à Raquel Rachid e a Luís Henrique Gonçalves pelas leituras críticas e sugestões de melhoria apresentadas neste trabalho.

Histórico do artigo: submetido: 30 maio 2021 | aceito: 25 ago. 2021 | publicado: 10 nov. 2021.

Apresentação anterior: não houve.

Licença CC BY-NC atribuição não comercial. Com essa licença é permitido acessar, baixar (*download*), copiar, imprimir, compartilhar, reutilizar e distribuir os artigos, desde que para uso não comercial e com a citação da fonte, conferindo os devidos créditos de autoria e menção à Reciis. Nesses casos, nenhuma permissão é necessária por parte dos autores ou dos editores.

INTRODUÇÃO

Atualmente a saúde digital vem sendo amplamente discutida nos fóruns acadêmicos, nas políticas públicas e nas organizações de saúde. Recentemente Lupton (2014) definiu a saúde digital como uma ampla faixa de tecnologias direcionadas a entregar cuidados de saúde, prover informação para as pessoas leigas e ajudá-las a compartilhar suas experiências de saúde e doença. Topol (2019) argumenta que o uso de grandes bases de dados (*big data*) e de inteligência artificial na saúde pode estabelecer as fundações para uma medicina de alto desempenho. A Organização Mundial da Saúde, com objetivo de acelerar a digitalização dos sistemas de saúde, produziu um guia para elaboração de estratégias nacionais de saúde digital (OMS; ITU, 2012). No Brasil, o Ministério da Saúde aderiu a essa proposta e elaborou uma Estratégia de Saúde Digital para o Brasil (BRASIL, 2020).

Não obstante as proposições citadas, os primeiros trabalhos referentes à aplicação de computadores na medicina remontam à década de 1950 em publicações nas áreas de biofísica, bioengenharia e eletrônica biomédica (COLLEN, 1986). Ainda nos anos 1980, Van Bommel (1984) definiu a informática médica como aspectos teóricos e práticos do processamento e da comunicação da informação baseados no conhecimento e nas experiências derivados dos processos em medicina e nos cuidados de saúde. Desse modo, o campo da informática médica trata de assuntos diversos da tecnologia e dos processos de saúde: sistemas de informação hospitalares, sistemas de suporte à decisão clínica, interoperabilidade de informações médicas, armazenamento de imagens médicas, telessaúde, registros eletrônicos de pacientes e mais recentemente tópicos como coleta de grandes volumes de dados a partir de *smartphones*, dispositivos vestíveis (*wearables*), redes sociais, armazenamento e computação em nuvem e a aplicação de inteligência artificial para a descoberta de padrões e para análise de dados médicos (imagens, textos, sinais vitais etc.).

Assim, podemos supor que a saúde digital em muito se inspira nos desenvolvimentos anteriores da informática médica; isto é, esse assunto deriva de décadas de construção de conhecimentos que remontam

à informática médica. Sendo assim, faz-se necessária uma análise mais rigorosa sobre como o campo de conhecimento da informática médica se desenvolveu ao longo do tempo.

A ciência da informação traz embasamento para, entre outras coisas, investigar os mecanismos de produção, representação, disseminação e avaliação da informação (CURTY; DELBIANCO, 2020), sendo que os estudos métricos da informação se dedicam à identificação e à avaliação da informação, ao seu alcance, à sua influência e ao seu impacto. Entre esses estudos métricos, a bibliometria trata de uma disciplina relacionada aos aspectos da comunicação de trabalhos científicos, examinando impacto, autoria, publicações, citações e conteúdos (HADDOW, 2018). Por meio dela, aplicam-se métodos quantitativos a um corpo da literatura (aqui definida como qualquer informação registrada) para explorar padrões de comunicação, tendências e redes de colaboração ou de coocorrência presentes naquela literatura. Esse tipo de análise fornece a pesquisadores, bibliotecários e administradores respostas sintéticas valiosas para seus estudos.

Este trabalho, portanto, utilizou técnicas e ferramentas bibliométricas para responder às seguintes questões: como o campo da informática médica evoluiu em relação a outras áreas científicas ao longo dos anos? Como foi a evolução geográfica desse campo? É possível identificar tendências históricas baseadas nos tópicos de pesquisa?

Desse modo, o trabalho tem como objetivo primário descrever, a partir da análise da literatura científica da informática médica, como este corpo de conhecimento tem evoluído e se transformado ao longo das décadas. O objetivo secundário é desenvolver processos bibliométricos para a avaliação da pesquisa, abordando a categoria ‘informática médica’, na Web of Science, no período de 1990 em diante. Com isso, para além de apresentarmos alguns desafios, suposições e limitações associados à pesquisa bibliométrica, apresentamos um estudo de caso real, no qual partimos de questões de pesquisa em nível abstrato mais alto e as respondemos por meio de diversas técnicas bibliométricas.

A compreensão de importantes indicadores – de como projetar estudos bibliométricos e usar ferramentas apropriadas – fornecerá aos iniciantes nesse campo de pesquisa o conhecimento e as habilidades para conduzir este tipo de pesquisa. Neste trabalho, detalhamos como utilizar técnicas bibliométricas por meio de *softwares* que implementam o ‘estado da arte’ nessas técnicas. Com isso, buscamos descrever metodologicamente a aplicação de análises avançadas de técnicas bibliométricas aplicadas ao estudo de um campo relacionado às ciências da saúde.

Para alcançar seus objetivos, este trabalho se organiza da seguinte forma: após esta introdução, a seção de referencial teórico apresenta conceitos de informática médica, saúde digital e bibliometria, bem como descreve ferramentas para coleta e análise de dados bibliográficos. A seção de metodologia descreve os procedimentos para coleta e análise dos dados em cinco etapas: desenho do estudo, coleta dos dados, análise dos dados, visualização dos dados e interpretação dos resultados. A seção de resultados apresenta efetivamente os resultados obtidos, após a coleta e o tratamento dos dados, principalmente por meio de contagem de frequências e gráficos. A seção de discussão realiza interpretações para a dinâmica científica do campo da informática médica, bem como elabora explicações para as mudanças observadas no campo. Por fim, as conclusões sintetizam os achados do trabalho, registram as limitações e indicam os possíveis estudos futuros.

REFERENCIAL TEÓRICO

Para consecução dos seus objetivos, este trabalho baseia-se em dois conjuntos de conhecimento: os estudos sobre informática médica e os estudos de bibliometria. Nesta seção apresentaremos esses corpos de conhecimento e explicaremos como esses conhecimentos serão mobilizados para analisar a evolução da informática médica por meio das técnicas de bibliometria.

Da informática médica à saúde digital

Esforços para estabelecer um campo de pesquisa específico levaram médicos e engenheiros a abordar aspectos científicos relativos às engenharias e à educação para constituir o campo da informática médica na década de 1960. Já na década de 1970 as primeiras conferências foram organizadas.

Como argumentou Collen (1986), as primeiras referências ao uso de computadores em medicina surgem na década de 1950 em publicações das áreas de biofísica, bioengenharia e eletrônica biomédica. Não havia, então, um termo consensual que delimitasse esse campo de aplicação específico. Foi somente nos anos 1970 que se tornou clara a necessidade de se definir um ‘rótulo’ que abarcasse esse campo e que fosse capaz de abranger suas distintas interfaces – medicina, pesquisa científica, engenharia e desenvolvimento tecnológico. A escolha recaiu sobre o termo ‘*medical informatics*’ – formado a partir do termo francês ‘*informatique*’ –, que passa a nomear o Comitê Técnico N° 4 da International Federation for Information Processing (IFIP). O termo tornou-se aceito internacionalmente em consequência de sua adoção pela IFIP na organização de seus congressos trienais sobre ‘informática médica’ (MEDINFO).

O primeiro MEDINFO, organizado pelo Comitê Técnico N° 4 da IFIP, ocorreu em agosto de 1974 em Estocolmo. Nos Estados Unidos, embora inicialmente tenha havido resistência à incorporação do termo, prevaleceu o padrão internacionalizado, que se consolida com a criação, em 1989, da American Medical Informatics Association (AMIA), a partir da fusão de duas organizações prévias – a American Association for Medical Systems and Informatics (criada em 1981) e o American College of Medical Informatics (criada em 1984).

A partir de então, o termo foi definido como “a aplicação de tecnologia de computadores a todos os campos da medicina – cuidado, ensino e pesquisa médica” (COLLEN, 1986, p. 779, tradução nossa). De acordo com Van Bommel (1984), a informática médica compreende os aspectos teóricos e práticos do processamento e da comunicação da informação baseados no conhecimento e nas experiências derivados dos processos em medicina e dos cuidados de saúde. Esse campo do conhecimento trata de assuntos diversos da tecnologia e dos processos de saúde; desde os sistemas de informação hospitalares, de suporte à decisão clínica – incluindo interoperabilidade de informações médicas, armazenamento de imagens médicas, telessaúde, registros eletrônicos de pacientes – até, mais recentemente, os tópicos de coleta de grandes volumes de dados a partir de *smartphones*, dispositivos vestíveis (*wearables*), redes sociais, computação em nuvem e a aplicação de inteligência artificial para a descoberta de padrões e para a análise de dados médicos (imagens, textos, sinais vitais etc.). Morris e McCain (1998) identificaram cinco grandes subáreas de pesquisa na informática médica: i) informática médica em geral; ii) educação médica; iii) informação para a tomada de decisão médica; iv) computação biomédica; e v) computação em engenharia biomédica.

Mais recentemente observa-se o surgimento da saúde digital, que a partir do uso de tecnologias digitais, tais como *big data*, computação em nuvem, dispositivos móveis, internet das coisas e inteligência artificial, propõe uma mudança paradigmática na área da saúde. Conforme citado, Topol (2019) argumenta que *softwares* baseados em *big data* e inteligência artificial irão ver e fazer coisas que não são humanamente possíveis. Com isso, é possível estabelecer as fundações para uma medicina de alto desempenho, baseada em dados, e que nos levará para além das somas das partes de humanos e máquinas.

Lupton (2014) reconhece o potencial da saúde digital, mas ao mesmo tempo pondera que se faz necessário investigar os sentidos e seus impactos nos conceitos de incorporação humana, subjetividade, relações sociais e instituições sociais. Nessa mesma linha, a Associação Brasileira de Saúde Coletiva (Abrasco), por meio de seu 3º Plano Diretor para o Desenvolvimento da Informação e Tecnologia de Informação em Saúde (2020), defende que “saberes e práticas de ITIS são determinados pelo contexto histórico, político, social, econômico e científico em que são gerados e desenvolvidos, constituindo-se em espaços plenos de relações

de poder e disputas de interesses” (p. 10). Lutpon (2014) sugere, portanto, ser primordial investigar e identificar as bases sociais, culturais e políticas da saúde digital.

O impacto potencial das tecnologias digitais no setor da saúde é grande e vem sendo apontado na literatura técnica. Ele inclui transformações tanto no setor assistencial quanto no industrial e numa inovação em larga escala de produtos e processos. Alcança também os arranjos institucionais que organizam e coordenam a prestação de serviços em diversos níveis, incluindo as ações de vigilância, promoção e prevenção, bem como as ações regulatórias e de gestão ao nível dos sistemas nacionais de saúde. Altera igualmente as formas através das quais os usuários se relacionam com os serviços e as organizações do setor, bem como as formas de produção, circulação e utilização de informações relacionadas à saúde.

A fim de contribuir com esse debate, o trabalho se engaja nas proposições de Lupton e descreve como o campo da informática médica passou por diversas mudanças nas últimas cinco décadas, chegando ao que conhecemos como saúde digital. Para tanto, utiliza-se de técnicas de bibliometria que serão descritas na próxima seção.

Bibliometria

A ciência da informação traz embasamento para, entre outras coisas, investigar os mecanismos de produção, representação, disseminação e avaliação da informação (CURTY; DELBIANCO, 2020), sendo que os estudos métricos da informação se dedicam à identificação e à avaliação da informação, no que diz respeito a seu alcance, sua influência e seu impacto.

Os estudos métricos da informação utilizam diversos aportes teóricos derivados de diferentes áreas de conhecimento, incluindo leis bibliométricas como: a Lei da Produtividade Científica (LOTKA, 1926), que descreve a frequência de publicação de autores em qualquer área; a Lei da Dispersão Bibliográfica (BRADFORD, 1934), similar à Distribuição de Pareto; e a Lei de Zipf (ZIPF, 1949), que rege a dimensão, importância ou frequência dos elementos de uma lista ordenada. Abordagens mais recentes foram criadas, a partir da influência das tecnologias da informação e da comunicação – em especial a *web*, que modificou as relações de produção, compartilhamento, acesso e uso da informação científica (CURTY; DELBIANCO, 2020) –, ao investigar aspectos quantitativos em relação a domínios, páginas *web* e *sites*.

O termo ‘bibliometria’ surge em 1969 com Alan Pritchard (PRITCHARD, 1969), que argumentou que esse conceito seria mais apropriado que o de ‘bibliografia estatística’, usado até aquele momento. O autor confere destaque para esse neologismo, que se relaciona com outros neologismos mais bem-aceitos pela comunidade científica, como os casos de econometria, biometria e cientometria. Curty e Delbianco (2020) identificam oito tipos de abordagens métricas mais bem consolidadas na literatura científica: altmetria, arquivometria, bibliometria, cibermetria, cientometria, informetria, patentometria e webometria. Algumas dessas abordagens mais recentes buscam informações complementares externas à produção em si, como a altmetria, que mede a atenção *on-line* que artigos científicos recebem por parte de diferentes redes sociais e recursos *web*. A bibliometria delimita-se em relação às outras abordagens ao usar como objetos de estudo as produções bibliográficas e foi uma das primeiras a ser criada, fazendo com que ainda haja confusão sobre as diferenciações entre cada abordagem (CURTY; DELBIANCO, 2020). A bibliometria ainda é usada para estudar a comunicação e a influência no escopo estritamente acadêmico dentro de um campo de pesquisa.

Originada da sociologia positivista, a bibliometria tem desenvolvido suas próprias leis e técnicas de mensuração. O fator de impacto e o índice Hirsch (ou índice H) são exemplos de indicadores aplicados a estudos bibliométricos.

Diferentemente da tradicional análise textual e da pesquisa histórica, a bibliometria busca quantificar padrões e tendências presentes nas características da literatura de interesse (HADDOW, 2018). A análise bibliométrica usa informações bibliográficas (títulos, autores, datas, referências, textos etc.) que descrevem

itens publicados na literatura científica para medir e investigar diferentes aspectos de estudo de um campo específico – fornecendo perspectivas sintéticas deste campo.

Análises bibliométricas da informática médica

Os primeiros estudos bibliométricos relacionados à informática médica datam da década de 1990. Morris e McCain (1998) analisaram 20 periódicos da área e aplicaram análise de cocitação para compreender como eles se relacionavam entre si. A técnica de cocitação também foi aplicada em Raghupathi e Nerur (2010) para entender as relações entre autores, nos anos de 1998 a 2006, bem como para identificar subáreas emergentes relacionadas à gestão e à internet. Liang (2010) investigou a informática em saúde usando a base de dados Web of Science em busca de autores, instituições, periódicos e artigos mais citados, além da historiografia das 30 publicações mais citadas.

Spreckelsen *et al.* (2011) analisaram métricas de periódicos, como fator de impacto, índice Hirsch, e índices relacionados aos autores – por exemplo, produtividade e impacto. Serenko *et al.* (2017) desenvolveram um *ranking* de periódicos acadêmicos de saúde eletrônica (*e-health*) baseado em uma *survey*, com a opinião dos pesquisadores da área de gestão e de prática clínica. Nadri *et al.* (2017) selecionaram os 100 artigos mais citados em informática médica de acordo com a plataforma Web of Science, identificando quais os tópicos mais comuns.

O conteúdo dos artigos completos foi analisado em outros trabalhos. Schuemie *et al.* (2009) extraíram variáveis textuais de títulos e resumos de artigos publicados em 20 periódicos categorizados como informática médica no Medline entre 1993 e 2008. Mehta *et al.* (2019) aplicaram revisão sistemática da literatura para identificar trabalhos que adotaram inteligência artificial na saúde. Gu *et al.* (2019) analisaram mais de três mil artigos da Web of Science entre 1992 e 2017, explorando palavras-chave em torno de tecnologias da informação, gestão da saúde e serviços de saúde eletrônica.

Outros estudos adotaram uma abordagem mista, usando indicadores estruturais e de conteúdo, buscando identificar agrupamentos de autores e instituições parceiras e apontar convergências nos interesses de pesquisa (LIANG, 2010; SWEILEH *et al.*, 2017; VOSNER; ZELEZNIK. KOKOL, 2019; GU *et al.*, 2019).

Neste artigo, buscamos trazer uma contribuição metodológica à pesquisa em informação em saúde, ao explorar diferentes indicadores bibliométricos de produções na área de informática médica, por meio de ferramentas que implementam as principais técnicas desenvolvidas na área. Demonstramos como elas podem auxiliar as pesquisas nas áreas relacionadas à informação em saúde e ajudar na compreensão de um campo e de sua evolução histórica e geográfica.

Ferramentas de bibliometria

Neste trabalho, duas ferramentas de análise bibliométrica com desenvolvimento recente foram empregadas: [Bibliometrix](#) (ARIA; CUCCURULLO, 2017) e [CorText](#). Ambos os *softwares* são gratuitos, estão em constante desenvolvimento e contam com uma comunidade ativa de usuários. O CorText foi concebido para ser uma ferramenta de apoio à pesquisa em métodos digitais aplicados às ciências sociais. Já o Bibliometrix foi desenvolvido a partir de referencial teórico da ciência da informação, trazendo as leis e métricas mais comuns desse campo de conhecimento. É importante considerar essa diferença na concepção de ambos, já que se trata de elemento que reflete nas funcionalidades disponíveis de cada ferramenta e no tipo de respostas que elas podem oferecer.

O Bibliometrix estrutura suas funcionalidades em torno de duas dimensões: o nível de análise e as estruturas de conhecimento. No nível de análise, são calculadas métricas e geradas visualizações para as seguintes unidades: fontes, autores e documentos (Quadro 1). Quanto às estruturas de conhecimento,

são abordadas as dimensões: conceitual (busca identificar tópicos, evolução temática), intelectual (redes de citações de fontes, artigos ou autores) e social (redes de colaboração entre autores, instituições e países) (Quadro 2). Essa divisão conceitual é espelhada no menu disponibilizado pela ferramenta e seu detalhamento pode ser encontrado na documentação *on-line*.

Quadro 1 – Funcionalidades do Bibliometrix, conforme o nível de análise

Nível de análise	Métricas
Fontes	Lei de Bradford Índice H Dinâmica das fontes Fontes mais relevantes
Autores	Autores mais relevantes Produção anual por autor Lei de Lotka Índice H Afiliações mais relevantes Países
Documentos	Documentos mais citados localmente e globalmente Referências citadas Dinâmica das palavras utilizadas

Fonte: adaptado a partir de Aria (c2016).

Quadro 2 – Funcionalidades do Bibliometrix, conforme as estruturas de conhecimento

Estrutura	Técnica bibliométrica	Unidade de análise	Técnicas estatísticas
Conceitual	Coocorrência de palavras	Título Resumo Palavras-chave Documento completo	Análise de redes Análise fatorial Mapeamento temático Modelagem de tópicos
Intelectual	Citação e cocitação	Referências Autores Periódicos	Análise de redes Historiografia
Social	Colaboração	Autores Instituições Periódicos	Redes de colaboração

Fonte: adaptado a partir do CorText Manager Documentation ([2015?]).

Trazendo outra perspectiva, o CorText é mais flexível no sentido de dar suporte a outros tipos de arquivos tabulados que não os estruturados, conforme os padrões das grandes bases científicas. Com isso, traz funcionalidades mais genéricas para processamento de texto, análises temporais, geográficas e de correlação entre as variáveis dos registros. O Quadro 3 sintetiza as suas principais funcionalidades.

Quadro 3 – Funcionalidades do CorText de acordo com suas categorias de análise

Categoria	Finalidade	Funcionalidades
Tratamento do corpus	Pré-processa os registros originais, de modo a prepará-los para as outras funcionalidades	Decomposição do <i>corpus</i> Subseleção de registros Categorização de dados numéricos Exploração do <i>corpus</i> Palavras em formato Word2vec Renomeação ou combinação de variáveis Análise de contraste
Processamento do texto	Cria listas de termos sinônimos, reconhece automaticamente as entidades nomeadas	Extração de termos Indexação dos termos no <i>corpus</i> Lista de entidades Indexação de lista de entidades Reconhecimento de entidades nomeadas Análise de sentimentos
Temporal	Extrai características dos registros do <i>corpus</i> a partir de uma ou mais variáveis temporais presentes nos metadados	Demografia Detector de períodos Fatiamento de períodos Análise de épocas
Espacial	Geolocaliza os registros de acordo com uma variável dos metadados	Geocodificação Exploração geoespacial
Análise	Apresenta técnicas matemáticas e estatísticas para explorar o relacionamento das palavras e os tópicos subjacentes aos registros	Mapeamento de redes Análise estrutural de redes Análise de correspondência Matriz de contingência Modelagem de tópicos Perfilamento
Método Sashimi	Combina modelagem estatística e mapas interativos para oferecer navegação hierárquica do <i>corpus</i> e da relação entre termos, documentos e metadados	Preparação do <i>corpus</i> Modelagem de tópicos do domínio Mapeamento de tópicos do domínio Dimensionamento em cadeia Mapa encadeado do domínio

Fonte: adaptado a partir do CorText Manager Documentation ([2015?]).

Neste trabalho, utilizamos as funcionalidades de ambas as ferramentas, que entendemos complementares para oferecer respostas de cunho bibliométrico a um *corpus*. Em nosso *corpus* de informática médica, que contém mais de 100 mil registros, ambas as ferramentas oferecem diferentes abordagens para as questões de pesquisa que estão postas – o que auxilia na compreensão da evolução dessa área de conhecimento.

METODOLOGIA

Neste trabalho, exploramos a produção científica da área de informática médica. A partir do cenário histórico exposto na introdução, focamos este artigo na análise bibliométrica dos trabalhos em informática médica desenvolvidos a partir da década de 1990. Além da consolidação da área nesta década, outro fator determinante para essa escolha foi a disponibilidade de palavras-chave a partir dessa data, oferecendo material para a análise dos temas usados nas pesquisas da área.

A Figura 1 demonstra os passos básicos para um estudo bibliométrico (ZUPIC; CATER, 2015). Este processo foi obtido a partir da combinação com outras metodologias de estudo criadas previamente, sintetizada e simplificada para esse formato.



Figura 1 – Metodologia básica para estudos bibliométricos
Fonte: adaptada de Zupic e Cater (2015).

Na fase de desenho do estudo, são definidas as perguntas de pesquisa a serem respondidas a partir de consulta à literatura e da aplicação das técnicas bibliométricas. A escolha das perguntas influencia quais funcionalidades estarão disponíveis ou quais técnicas deverão ser aplicadas. As unidades de análise também deverão ser definidas de acordo com as perguntas (documento, autores, termos-chave, títulos, afiliação dos autores etc.).

Neste artigo foram estabelecidas as questões de pesquisa a serem respondidas sobre a evolução da área de informática médica, por meio da abordagem bibliométrica (Quadro 4). As questões de pesquisa postas aqui não são questões diretamente respondidas por métricas de produção; no entanto, aproximam-se mais dos interesses de possíveis leitores acerca do assunto abordado. Elas foram desmembradas em métricas subjacentes para serem respondidas sob diferentes perspectivas.

Quadro 4 – Questões de pesquisa do estudo de caso

- QP1) Como o campo da informática médica evoluiu ao longo dos anos?
- QP2) Como foi a evolução geográfica da área?
- QP3) É possível identificar tendências históricas baseadas nos tópicos de pesquisa?

Fonte: elaborado pelos autores.

Na coleta dos dados, são aplicadas técnicas de revisão da literatura como as de mapeamento ou de revisão sistemática. De forma geral, são escolhidas as bases de dados relevantes para o campo de conhecimento sob estudo e os termos a serem pesquisados. Bases multidisciplinares como a Web of Science e Scopus geralmente são incluídas. Bases específicas para as ciências da saúde também são incluídas, como é o caso da PubMed e da Cochrane, com os resultados sendo posteriormente combinados, e eliminando-se possíveis itens duplicados. Os termos usados para as buscas também são definidos e refinados. É uma boa prática expandir os termos com descritores relacionados (sinônimos, hipônimos e hiperônimos, por exemplo), consultados a partir de taxonomias reconhecidas, como a [MeSH](#) (Medical Subject Headings) e a brasileira [DeCS](#) (Descritores em Ciências da Saúde). Em seguida, é necessário filtrar por possíveis categorias, como tipo de publicação, idioma, período ou veículo. Cada base de dados possui suas regras para fazer a busca, de modo que a *string* de busca com os termos deve ser adaptada para cada uma delas. Uma vez executada a busca, é necessário baixar os metadados dos resultados, para que seja possível extrair os indicadores bibliométricos de interesse.

As unidades de análise adotadas foram os documentos e seus conjuntos de palavras-chave. Foi feita a busca na base Web of Science, filtrando pela categoria ‘*medical informatics*’ e por trabalhos em inglês publicados em periódicos, conferências e revisões. No total, foram recuperados 100.315 registros, que

datavam de 1961 até o fim de 2020. Os metadados dos registros foram coletados a partir da plataforma *web*, sendo baixados em lotes de 500 registros por vez.

Estes metadados foram importados nas ferramentas Bibliometrix e CorText, usadas neste trabalho para aferir os indicadores bibliométricos de interesse para o presente estudo.

O *software* CorText é uma plataforma *web*, não sendo necessária a instalação de outros programas para que seus usuários possam utilizá-lo. O Bibliometrix é um pacote escrito na [linguagem R](#) – muito comum na comunidade de estatística e de computação. Como ele é executado localmente na máquina do usuário, é necessária a instalação do *kit* de desenvolvimento da linguagem R, que varia conforme o sistema operacional. Após o *download* e a instalação da linguagem R, são necessários os seguintes comandos: `install.packages("bibliometrix")`, para que o pacote Bibliometrix seja instalado; `library(bibliometrix)`, para carregar a biblioteca com as funcionalidades; e, em seguida, `biblioshiny()`, para que a interface com o usuário apareça no navegador do computador do usuário. Uma das vantagens do Bibliometrix é ser uma ferramenta que não necessita de conhecimentos de programação para ser utilizada. No entanto, todas as funcionalidades aqui mostradas podem ser executadas e personalizadas mediante a execução de códigos escritos na linguagem R, que podem ser consultados na documentação do pacote.

Na fase análise dos dados, são calculadas as métricas de interesse, de acordo com as questões da pesquisa. As ferramentas aqui ilustradas oferecem suporte particular a essa fase, ao prover uma grande quantidade de funcionalidades que atende a esse momento da pesquisa, sem ser necessário conhecimento computacional avançado. Exemplos dessas funcionalidades serão mostrados na seção a seguir.

Em ambas as plataformas, é necessário importar os metadados dos registros para que eles possam ser processados. Cada uma delas trata de forma diferente essa ação. No CorText, é necessário cadastrar um usuário e criar um projeto. Deve-se criar um único arquivo compactado (com extensão *zip*, obrigatoriamente) que contenha todos os registros coletados. Esses registros podem estar organizados em diferentes estruturas de pastas – o *software* fará com que todos sejam considerados um mesmo *corpus* dentro do projeto. O *upload* é feito para a plataforma e, a partir desse momento, os dados ficam disponíveis para a aplicação das técnicas bibliométricas. No Bibliometrix, é possível importar os dados originais de diferentes bases, por meio de uma API (Interface de Programação de Aplicação) – apenas PubMed, no momento – ou por meio de um arquivo em formato interpretável pela linguagem R. O método usual é importar todos os arquivos em seus formatos originais e exportar apenas um, em formato de dados do R e, assim, facilitar seu uso posterior. As Figuras 2 e 3 ilustram as interfaces com os usuários de ambas as ferramentas.

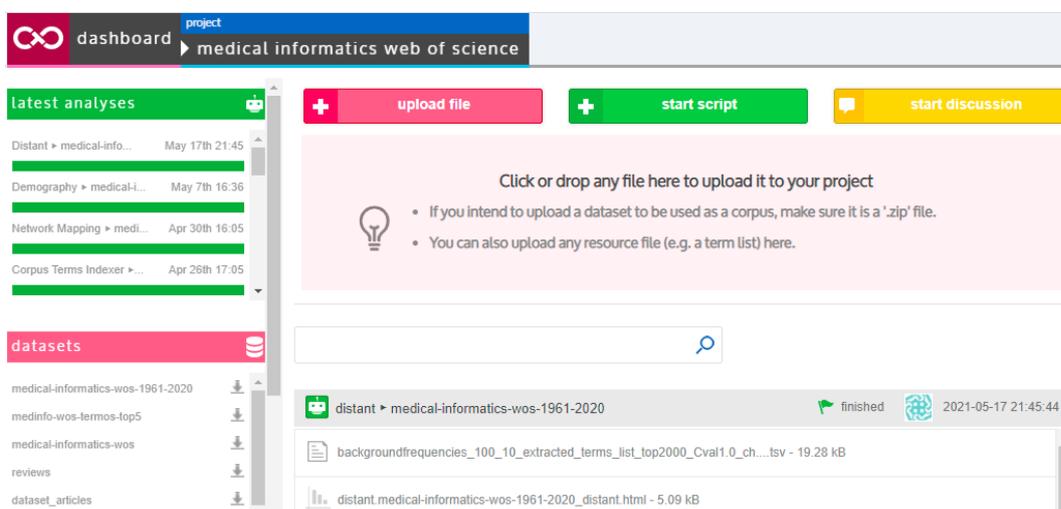


Figura 2 – Tela inicial do CorText

Fonte: print realizado pelos autores.

The screenshot shows the Bibliometrix web interface. On the left, there is a sidebar with the following sections:

- Import or Load:** Includes a dropdown for 'Please, choose what to do' (set to 'Load bibliometrix file(s)'), a 'Browse...' button, a file name 'bibliometrix_v3_regeix_line_breaks', an 'Upload complete' indicator, and a 'Start' button.
- Conversion results:** Shows 'Number of Documents 100315'.
- Export a bibliometrix file:** Includes a 'Save as:' dropdown menu.

On the right, there is a data table with the following columns: DOI, AU, AF, DE, ID, CR, and AB. The table contains three rows of data:

DOI	AU	AF	DE	ID	CR	AB
10.1002/ijoc.1097 0258.1997.1720.24.24.5987-AD- 9M458+310CO-2	GAIL MH	GAIL MH		MANTEL-HAENSEL ESTIMATOR; RELATIVE EFFICIENCY; STATISTICS; TESTS; NUMBER; MODELS	ANDERSON JA, 1972, BIOMETRIKA, V59, P19, DOI 10.1092/BIOMET/59.119ARMITAGE P, 1955, BIOMETRICS, V11, P391, DOI 10.2307/2301795, BIRCH HW, 1954, J RBY	NATHAN MANTELS MANY CONTRIBUTIONS TO EPIDEMIOLOGY GREW OUT OF HIS ROLE AS A STATISTICAL CONSULTANT AND COLLABORATOR. HIS ABILITIES TO UNDERSTAND SCIE
NA	GREENHOUSE SW	GREENHOUSE SW			GOLDIN A, 1953, J NATL CANCER I V13, P1443; MANTEL N, 1972, AM STAT, V27, P21, DOI 10.2307/2682206; MANTEL N, 1969, AM STAT, V23, P21; MANTEL N, 1966, J	NATHAN MANTEL SERVED AS A STATISTICIAN IN THE NATIONAL CANCER INSTITUTE OF THE NIH FOR MORE THAN 25 YEARS. DURING THIS PERIOD HE WAS A CONSULTANT TO M
10.1002/ijoc.1097 0258.1997.1720.24.24.5987-AD- 9M377+310CO-2	WALTER SOJADAD AR	WALTER SOJADAD, AR		RANDOMIZED CONTROLLED TRIALS; COLORECTAL CANCER; BREAST CANCER; SYSTEMATIC REVIEWS; VENOUS THROMBOSIS; PROSTATE CANCER; OVARIAN CANCER; DOWNS SYNDROME	ACKERMANN SF, 1994, J CANCER EDUC, V9, P96; ANJUM WM, 1996, FERTIL STERIL, V65, P1072, DOI 10.1016/S0015-0282(16)58220-4; AUSTIN SM, 1994, J AM MED INF	META-ANALYSIS HAS BECOME A POPULAR TECHNIQUE IN MANY AREAS OF BIOMEDICAL RESEARCH. WHILE THERE HAVE BEEN A NUMBER OF STUDIES AND COMMENTARIES ON THE

Figura 3 – Tela inicial do Bibliometrix

Fonte: print realizado pelos autores.

A fase de visualização dos dados também se beneficia das ferramentas aqui expostas, já que estas contêm funcionalidades de visualização dos dados de acordo com o tipo de análise selecionada. Muitas das técnicas bibliométricas apresentam a possibilidade de geração de gráficos e exploração interativa sobre os dados, proporcionando o exame de subcategorias de interesse de uma maneira mais detalhada (conhecido como *drill-down*).

Por fim, na fase de interpretação dos dados, o pesquisador discute os resultados encontrados, argumenta sobre as evidências encontradas e sobre os pressupostos teóricos das leis e medições obtidas. A partir dos resultados devem ser discutidos os padrões, as tendências e os pontos fora da curva encontrados. As limitações também devem ser consideradas, uma vez que a abordagem bibliométrica fornece uma visão geral sobre um *corpus* de conhecimento, sem buscar explicar as causas para sua formação e evolução.

RESULTADOS

Nesta seção, abordamos as análises e visualizações simultaneamente, uma vez que as ferramentas já geram visualizações dos dados, a partir dos cálculos das métricas bibliométricas.

Para a primeira questão de pesquisa, usamos estatísticas descritivas distribuídas anualmente pelo período considerado. A comparação é feita com o total de artigos na base Web of Science, sem o filtro *'medical informatics'*. Analisamos o crescimento absoluto e relativo nos artigos publicados nessa área de pesquisa. Tais estatísticas oferecem visões gerais sobre a estrutura do *corpus* em estudo. Os registros foram agregados por ano e exportados para o Excel para gerar a Figura 4.

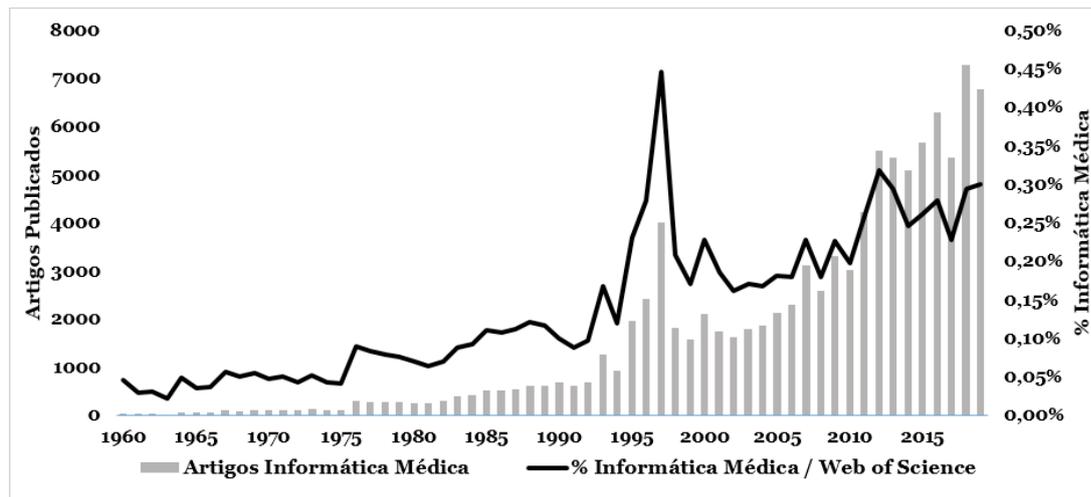


Figura 4 – Crescimento anual absoluto e relativo da área de informática médica em relação à ciência como um todo
Fonte: elaborada pelos autores.

A questão de pesquisa 2 busca fornecer uma visão sobre a abrangência internacional do campo ao longo dos anos. Para isso, utilizamos o Bibliometrix para identificar quais países contribuíram, em quatro períodos a saber: 1961-1989, 1990-2000, 2001-2010 e 2011-2020. O *software* o software calcula o número de artigos publicados por país presente, de acordo com a afiliação dos autores presente nos metadados. O Bibliometrix oferece a opção de produção científica por país (*Country Scientific Production*), dentro da área de autores (*Authors*). A Figura 5 ilustra o resultado, conforme os quatro períodos identificados, da esquerda para a direita e de cima para baixo. Quanto mais escura a cor azul, maior o número de artigos publicados por autores daquele país. Países em cinza não tiveram nenhuma publicação no período considerado.

Podemos notar que o primeiro período foi dominado por países do norte, em especial da América do Norte e da Europa Ocidental. No segundo período, novos países aparecerem, como China, Quênia e Austrália, muitos deles a partir de parcerias com países presentes no período anterior. A América do Sul começa a aumentar o número de países com publicação a partir do terceiro período. Já o quarto período apresenta a popularização em países da África.

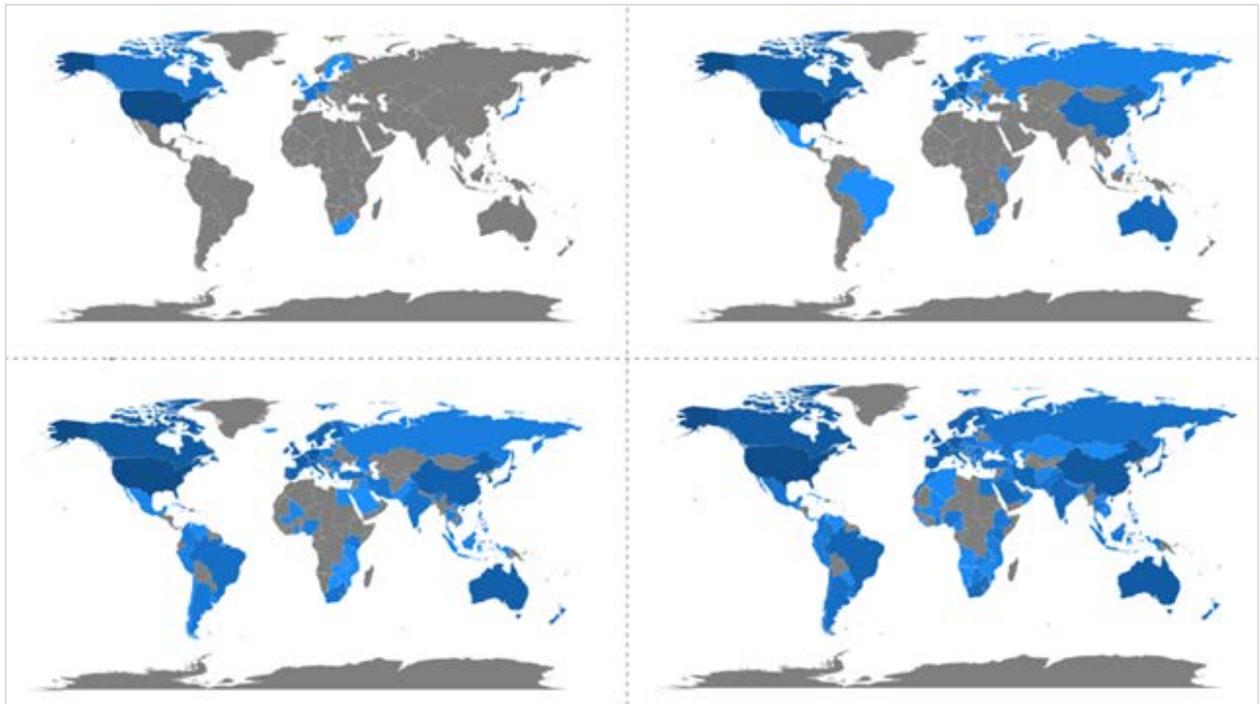


Figura 5 – Evolução geográfica da origem dos autores na área de informática médica
Fonte: elaborada pelos autores.

Para a terceira questão da pesquisa, utilizamos uma funcionalidade do CorText que calcula o grau de similaridade da produção de trabalhos em diferentes anos. Nesse caso, o *software* monta um vetor de palavras-chave para cada ano considerado. Esse vetor contém o número em que determinado termo aparece entre as palavras-chave daquele ano. Em seguida, é calculado o grau de similaridade para cada par de anos, por meio de um cálculo de similaridade entre vetores multidimensionais. O algoritmo então detecta quais recortes podem ser feitos, resultando em períodos com palavras-chave similares entre si e diferentes dos outros períodos. A Figura 6 ilustra o resultado da aplicação desse algoritmo. São identificados quatro períodos (1990-1994, 1995-2000, 2001-2010, 2011-2020).

Porém, para facilitar a análise, os dois primeiros períodos foram combinados. Assim, definimos três décadas de análise (1990-2000, 2001-2010, 2011-2020). Essa divisão permite analisar diferentes ciclos dentro de uma área de conhecimento, evidenciando o surgimento de eventos que impactaram a produção científica da área. No entanto, a bibliometria não determina quais são esses eventos, sendo uma limitação em estudos desse tipo.

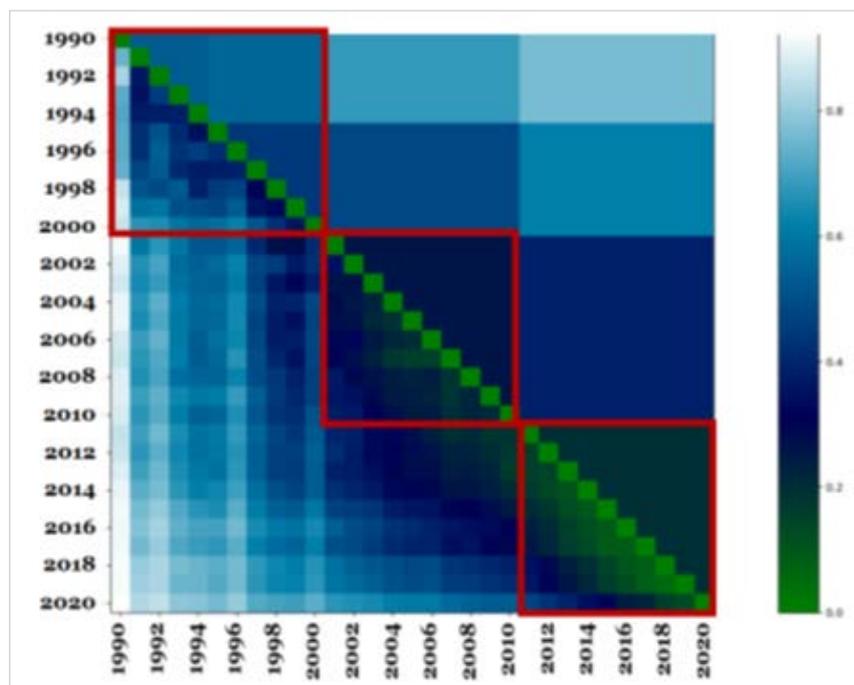


Figura 6 – Períodos detectados a partir do grau de semelhança de palavras-chave utilizadas a cada ano
Fonte: elaborada pelos autores.

Utilizamos a divisão das três décadas para segmentar o *corpus* no Bibliometrix. A ferramenta conta com a opção de filtragem por ano, facilitando a análise desses *subcorpora*. Nesse estudo de caso, caracterizamos cada um dos períodos buscando entender quais foram as variações de assuntos estudados entre eles. Para isso, usamos novamente o Bibliometrix para fazer a contagem das palavras-chave mais comuns a cada período, o que oferece *insights* sobre as transformações ocorridas nesse campo de conhecimento entre os dois períodos detectados. A Figura 7 ilustra a comparação das 10 palavras-chave mais comuns em dois desses períodos.

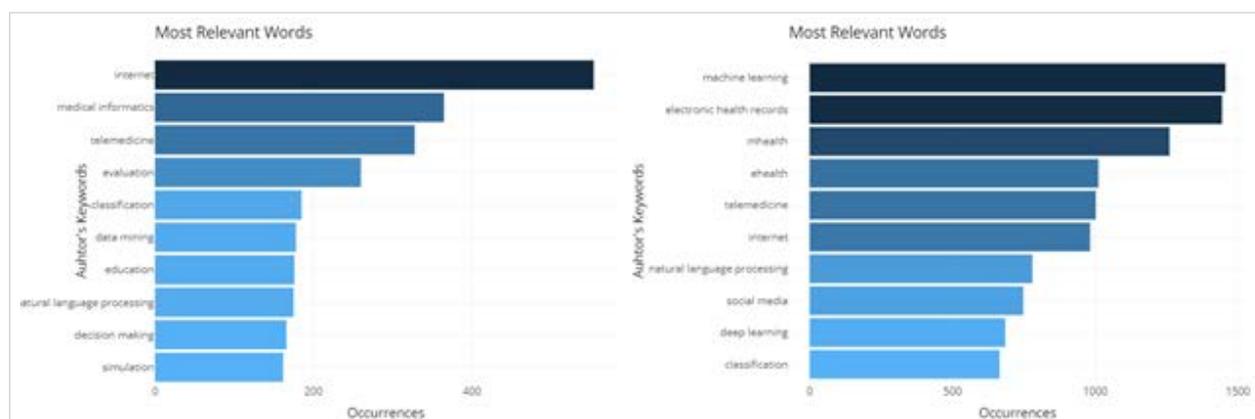


Figura 7 – Palavras-chave mais comuns em dois dos períodos detectados: 2000-2010 (à esquerda) e 2011-2020 (à direita)

Fonte: elaborada pelos autores.

Nesse estudo, utilizamos as palavras-chave extraídas para criar categorias sobre sua finalidade, focando em dois grupos: i) os rótulos usados pela própria área para se definir; e ii) as tecnologias e as aplicações de informática médica. O Quadro 5 sintetiza a frequência dos rótulos em cada década e as principais tecnologias e aplicações de informática médica no período correspondente.

Quadro 5 – Síntese das palavras-chave identificadas em cada período e suas respectivas frequências

	1990-2000	2001-2010	2011-2020
Rótulos	<i>Medical informatics</i> (106) <i>Health informatics</i> (46)	<i>Medical informatics</i> (401) <i>Health informatics</i> (194) <i>e-health</i> (175)	<i>e-health</i> (1.326) <i>Medical informatics</i> (702) <i>Health IT</i> (444) <i>Health informatics</i> (443) <i>Digital health</i> (248)
Tecnologias e aplicações	<i>Neural networks</i> (219) <i>Expert systems</i> (156) <i>PACS</i> (129) <i>Hospital information systems</i> (102) <i>Telemedicine</i> (94) <i>Internet</i> (93) <i>Simulation</i> (84) <i>Image processing</i> (80) <i>Clinical decision support</i> (78) <i>Electronic patient record</i> (41)	<i>Internet</i> (530) <i>Telemedicine</i> (320) <i>Electronic health records</i> (292) <i>Clinical decision support</i> (282) <i>Neural networks</i> (241) <i>Ontology</i> (204) <i>Natural language processing</i> (180) <i>Data mining</i> (173) <i>Hospital information systems</i> (168)	<i>Electronic health Records</i> (2.389) <i>Machine learning</i> (2.141) <i>Mobile health</i> (2.035) <i>Smartphone</i> (1.190) <i>Clinical decision support</i> (1.158) <i>Telemedicine</i> (1.015) <i>Internet</i> (982) <i>Social media</i> (949) <i>Natural language Processing</i> (827) <i>Telehealth</i> (305)

Fonte: elaborado pelos autores.

DISCUSSÃO

Apresentamos neste trabalho um estudo métrico da informação científica em saúde, mais precisamente na subárea da informática médica. Esse recorte foi adotado por seu escopo alinhado ao interesse desta pesquisa e pela conveniência da existência de um código descritor presente na base de dados da Web of Science que indexa os periódicos e as conferências, conforme esse assunto. Com isso, esperou-se maximizar a quantidade de verdadeiros positivos retornados pela base de dados. Por outro lado, essa opção pode ter criado falsos negativos, ou seja, artigos de interesse que não foram computados na criação do *corpus*.

Outra limitação diz respeito ao uso da Web of Science como única fonte de dados e *proxy* para a literatura científica. Essa base foi escolhida por abordar diferentes campos de conhecimento e compartilhar sobreposições com outras grandes bases. Argumentamos, entretanto, que a quantidade de falsos positivos ou o uso de apenas uma base de dados não tende a ameaçar as conclusões gerais aqui apresentadas. Embora não explique o porquê desses eventos, a abordagem bibliométrica forneceu o suporte para indicar ‘o que’ foi pesquisado, ‘como’ foi caracterizado, ‘quando’ ocorreram as grandes transformações, ‘onde’ se situavam as instituições em que as pesquisas foram conduzidas.

Comparada a outras áreas da ciência em geral – aqui representada pelo *proxy* do catálogo da Web of Science –, a informática médica demonstra um crescimento consistente ao longo das últimas seis décadas, em termos de artigos, autores, instituições, países, periódicos e assuntos. Em seus primeiros anos, foram publicados em torno de 50 artigos por ano, representando aproximadamente 0,05% de todos os artigos indexados na base de dados nesse período. No ano de 2020, foram publicados 7.098 artigos categorizados na área de informática médica, representando 0,30% entre todas as categorias indexadas na mesma base.

Em que pese o caráter circunstancial de sua adoção nos anos 1970, o termo ‘informática médica’ se impôs e vem sendo utilizado recorrentemente. Um exemplo é a própria Web of Science, em que ‘*medical informatics*’ é utilizado como rótulo tanto para indexação de itens bibliográficos quanto de periódicos. Os resultados da pesquisa aqui apresentada confirmam a resiliência da expressão ao longo do tempo (ver adiante), não obstante o surgimento de outros ‘rótulos’ em tempos mais recentes, como ‘*e-health*’ e, finalmente, ‘*digital health*’. Segundo a hipótese inicial que orientou nossa pesquisa, o surgimento desses

novos termos associa-se a mudanças tecnológicas e a aplicações correspondentes no campo da informática médica que demarcam transições temporais. Agregamos a essa hipótese inicial uma hipótese metodológica: as mudanças terminológicas no âmbito da literatura científica, indexada em bases de dados como a mencionada Web of Science, podem oferecer uma visão aproximativa das transformações do próprio campo de aplicação da ‘informática médica’ ao longo do tempo, uma vez que houvesse quantidade de registros suficiente para inferir padrões na base de dados.

Por meio da análise de palavras-chave, pudemos identificar três períodos distintos. Argumentamos que o campo de pesquisa da informática médica evoluiu por meio de ondas coerentes e diferentes entre si. Originalmente denominado informática médica, esse termo, a partir dos anos 1990, mudou para informática em saúde (*health informatics*), saúde eletrônica (*e-health*) e mais recentemente saúde digital (*digital health*). A análise das palavras-chave mostra que no início dos anos 1990 o campo era bastante heterogêneo em termos de assuntos pesquisados, contudo, a comunidade passou a convergir para alguns temas a partir da segunda metade da década de 1990. Essa convergência pode ser atribuída aos esforços para a definição de um vocabulário comum no campo, e entre seus resultados podemos mencionar a normatização do Registro Eletrônico de Saúde pelo National Institutes of Health e a posterior definição de um padrão ISO em 2005 (ISO, 2005).

Nos anos 2000, observa-se a expansão do campo para outras profissões da saúde, como a enfermagem. Assim, ganha força a nova denominação informática em saúde, que é mais abrangente que informática médica. Além disso, emergem os registros eletrônicos de saúde, junto com as aplicações de telemedicina.

Na última década, observa-se o estabelecimento do termo saúde eletrônica (*electronic health*, ou *e-health*), muito provavelmente uma apropriação do sucesso do comércio eletrônico, que se expandia no mesmo período. Nesses anos, também chama a atenção o vertiginoso crescimento da saúde móvel, em que se observa uma mudança, com a redução do uso de computadores pessoais em favor dos *smartphones* nas aplicações de saúde.

O termo saúde digital aparece somente em 2011, mas tem um crescimento expressivo a partir de 2017. Além disso, observa-se que a saúde digital tem sido adotada nos documentos de políticas públicas (OMS, 2020), bem como nas editoriais científicas, a exemplo da Lancet, que em 2019 lançou um periódico dedicado à saúde digital (The Lancet Digital Health). Essas evidências dos últimos 30 anos, aliadas aos dados mais recentes, nos levam a crer que na próxima década veremos a predominância do assunto *digital health* na literatura científica de informática médica.

Outras possibilidades de análises bibliométricas podem ser desenvolvidas, como o estudo das fontes ou da colaboração entre autores ou países. As questões de pesquisa e as suas respostas podem ser guiadas pelas informações tratadas neste artigo, sendo possível refinar e selecionar a ferramenta mais adequada para cada recorte desejado. Outras ferramentas estão disponíveis e podem complementar uma análise bibliométrica, como BibExcel (PERSSON; DANELL; SCHNEIDER, 2009), VOSviewer (VAN ECK; WALTMAN, 2007), HistCite (GARFIELD; SHER; TORPIE, 1964), BibTechMon (NOLL; FRÖHLICH; SCHIEBEL, 2002), [VantagePoint](#), entre outros – cada um deles trazendo funcionalidades similares às apresentadas, com diferentes graus de facilidade de uso e de sofisticação dos resultados.

É importante reconhecer as limitações e a complementaridade das ferramentas. A resposta a diferentes questões de pesquisas pode exigir diferentes métricas presentes em diferentes ferramentas, cabendo ao pesquisador conhecer as potencialidades e limitações de cada uma delas. Neste artigo demonstramos como atender metodologicamente a um cenário que exige diferentes perspectivas sobre a evolução de um campo de pesquisa.

CONCLUSÃO

Este trabalho analisou o campo de pesquisa da informática médica por meio da base de dados da Web of Science e, mediante técnicas bibliométricas, apresentou a evolução temporal da área nos últimos 30 anos, mostrando variações em relação aos termos usados para sua nomenclatura e as principais tecnologias desenvolvidas.

Buscamos analisar toda a produção científica dessa subárea dentro da Web of Science e, para isso, adotamos uma abordagem metodológica baseada na área de bibliometria, misturando abordagens qualitativas e quantitativas e suportadas por ferramentas que permitem essas análises de forma semiautomática. A presente investigação aborda a evolução da área de pesquisa da informática médica baseada em informações estruturais, sobretudo as palavras-chave e a filiação dos autores.

As ferramentas Bibliometrix e CorText foram usadas para fornecer suporte às questões de pesquisa, complementando-se em suas funcionalidades. O Bibliometrix necessita de algum conhecimento em instalação e configuração de programas para executá-lo; porém, traz prontamente as métricas mais comuns para estudos bibliométricos, utilizando os metadados dos artigos. Já o CorText é uma ferramenta mais genérica de processamento de texto que resolve ambiguidades, categoriza, cria *clusters* e identifica padrões temporais mais facilmente, sem ser necessária sua instalação.

A partir de nossos resultados, foi possível detectar quatro períodos cujos tópicos de pesquisa – representados pelas palavras-chave dos respectivos artigos – apresentaram características de acordo com o desenvolvimento tecnológico e a maturidade da área. Além disso, evidências dos últimos anos sugerem o surgimento de uma nova fase, a saúde digital, impulsionada por tecnologias de *smartphones*, armazenamento em nuvem e aprendizagem de máquina para o tratamento de problemas relacionados a diagnóstico precoce, análise de imagens médicas, atendimento remoto, autogestão da saúde, entre outros. Para uma visão complementar das características da evolução presente neste estudo metodológico, referir-se ao trabalho de Fornazin *et al.* (2021).

Esta pesquisa apresenta limitações inerentes à abordagem bibliométrica. A pesquisa bibliométrica varia das estatísticas descritivas de um corpo da literatura até modelagens sofisticadas de redes sociais, podendo conter milhares de registros. Problemas como taxas de participação ou de resposta, comuns em pesquisas de coleta de questionários, não ocorrem. Por outro lado, pesquisas bibliométricas apresentam limitações quanto às conclusões sobre o fenômeno, que nem sempre pode ser explicado pelas estatísticas geradas. Além de conhecer profundamente as ferramentas e métricas usadas na pesquisa bibliométrica, é necessário que o pesquisador saiba reconhecer as limitações dessa abordagem e se familiarize com o debate em torno do que pode ser medido.

Este é um projeto em desenvolvimento, e, como trabalhos futuros, espera-se o aprofundamento qualitativo sobre cada um dos períodos, investigando obras e autores representativos e buscando mapear seus desdobramentos e sua influência na área. Serão buscados, também, o aprofundamento teórico dos ciclos tecnológicos e o desenvolvimento de teoria da evolução tecnológica na área da saúde.

REFERÊNCIAS

ARIA, Massimo; CUCCURULLO, Corrado. Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, [s. l.], v. 11, n. 4, p. 959-975, 2017. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157717300500>. Acesso em: 22 set. 2021.

ARIA, Massimo. **Key Points**. Nápoles: Bibliometrix, c2016. Disponível em: <https://www.bibliometrix.org/Keypoints.html>. Acesso em: 22 set. 2021.

ASSOCIAÇÃO BRASILEIRA DE SAÚDE COLETIVA (ABRASCO). **3º Plano Diretor para o Desenvolvimento da Informação e Tecnologia de Informação em Saúde**. Rio de Janeiro: A Associação, 2020. Disponível em: <https://www.abrasco.org.br/site/gtinformacoesensaudepopulacao/wp-content/uploads/sites/13/2021/04/plad5.pdf>. Acesso em: 21 set. 2021.

BRADFORD, Samuel C. Sources of information on specific subjects. **Engineering: an Illustrated Weekly Journal**, Londres, n. 137, p. 85-86, 1934.

BRASIL. Ministério da Saúde. **Estratégia de Saúde Digital para o Brasil 2020-2028**. Brasília, DF: UNASUS, 2020. Disponível em: https://bvsmis.saude.gov.br/bvs/publicacoes/estrategia_saude_digital_Brasil.pdf. Acesso em: 17 ago. 2021.

COLLEN, Morris F. Origins of medical informatics. **Western Journal of Medicine**, São Francisco, v. 145, n. 6, p. 778-785, 1986. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1307150/>. Acesso em: 21 set. 2021.

CORTEXT Manager Documentation. Noisy-le-Grand: Inrae, [2015?]. Disponível em: <https://docs.cortext.net/>. Acesso em: 22 set. 2021.

CURTY, Renata Gonçalves; DELBIANCO, Natalia Rodrigues. As diferentes métricas dos estudos métricos da informação: evolução epistemológica, inter-relações e representações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 25, p. 1-21, 2020. DOI: <https://doi.org/10.5007/1518-2924.2020.e74593>. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/74593>. Acesso em: 10 ago. 2021.

FORNAZIN, Marcelo *et al.* From medical informatics to digital health: a bibliometric analysis of the research field. *In: AMERICAS' CONFERENCE ON INFORMATION SYSTEMS*, 9-13 ago. 2021. **Proceedings**. Atlanta: Association for Information Systems, 2021. Disponível em: https://aisel.aisnet.org/amcis2021/healthcare_it_sig_health/18. Acesso em: 17 ago. 2021.

GARFIELD, Eugene; SHER, Irving; TORPIE, Richard. **The use of citation data in writing the history of science**. Filadélfia: New American Library, 1964. Disponível em: www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf. Acesso em: 10 ago. 2021.

GU, Dongxiao *et al.* Visualizing the intellectual structure and evolution of electronic health and telemedicine research. **International Journal of Medical Informatics**, Shannon, v. 130, 2019. DOI: <https://doi.org/10.1016/j.ijmedinf.2019.08.007>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1386505618306774>. Acesso em: 10 ago. 2021.

HADDOW, Gaby. Bibliometric research. *In: WILLIAMSON, Kirsty; JOHANSON, Graeme (Ed.) Research methods: Information, systems, and contexts*. 2. ed. Victoria: Elsevier, 2018. p. 241-266.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO/TR 20514:2005**: Health informatics – Electronic health record – Definition, scope and context. Genebra: A Organização, 2005. Disponível em: <https://www.iso.org/standard/39525.html>. Acesso em: 18 ago. 2021.

LIANG, Hai-Ning. Overview of the health informatics research field: a bibliometric approach. *In: TAKEDA, Hiroshi (Ed.) IMIA/IFIP Joint Symposium on e-health*, Berlim: Springer, 2010. v. 335, p. 37-48. DOI: https://doi.org/10.1007/978-3-642-15515-4_5. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-15515-4_5. Acesso em: 10 ago. 2021.

LOTKA, Alfred. The frequency distribution of scientific productivity. **Journal of the Washington Academy of Sciences**, Washington, DC, v. 16, n. 12, p. 317-323, 1926. Disponível em: <https://www.jstor.org/stable/24529203?refreqid=excelsior%3A5dba49cb4030f0cc6c4873b1b3a69ea5>. Acesso em: 21 set. 2021.

LUPTON, Deborah. The commodification of patient opinion: the digital patient experience economy in the age of big data. **Sociology of Health & Illness**, Oxford, v. 36, n. 6, p. 856-869, 2014. DOI: <https://doi.org/10.1111/1467-9566.12109>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/1467-9566.12109>. Acesso em: 21 set. 2021.

MEHTA, Nishita; PANDIT, Anil; SHUKLA, Sharvari. Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. **Journal of Biomedical Informatics**, [s. l.], v. 100, p. 1-14, 2019. DOI: <https://doi.org/10.1016/j.jbi.2019.103311>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1532046419302308>. Acesso em: 10 ago. 2021.

MORRIS, Theodore; McCAIN, Katherine. The structure of medical informatics journal literature. **Journal of the American Medical Information Association**, Filadélfia, v. 5, n. 5, p. 448-466, 1998. DOI: <https://dx.doi.org/10.1136%2Fjamia.1998.0050448>. Disponível em: <https://academic.oup.com/jamia/article/5/5/448/792285>. Acesso em: 10 ago. 2021.

NADRI, Hamed *et al.* The top 100 articles in the medical informatics: a bibliometric analysis. **Journal of Medical Systems**, New York, v. 41, n. 10, 2017. DOI: <https://doi.org/10.1007/s10916-017-0794-4>. Disponível em: <https://link.springer.com/article/10.1007%2Fs10916-017-0794-4>. Acesso em: 10 ago. 2021.

NOLL, Margit; FRÖHLICH, Doris; SCHIEBEL, Edgar. Knowledge maps of knowledge management tools: information visualization with BibTechMon. In: KARAGIANNIS, Dimitris; REIMER, Ulrich (Ed.). **Practical aspects of knowledge management: 4th International Conference, PAKM 2002, Vienna, Áustria, December 2-3, Proceedings**. Berlin: Springer, 2002. p. 14-27.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS). **Global strategy on digital health 2020-2025**. Genebra: A Organização, 2020. Disponível em: <https://www.who.int/docs/default-source/documents/g4dhdad2a9f352b0445bafbc79ca799dce4d.pdf>. Acesso em: 17 ago. 2021.

ORGANIZAÇÃO MUNDIAL DA SAÚDE (OMS); INTERNATIONAL TELECOMMUNICATION UNION (ITU). **National e-health strategy toolkit**. Genebra: Organização Mundial da Saúde, 2012. Disponível em: <https://apps.who.int/iris/handle/10665/75211>. Acesso em: 17 ago. 2021.

PERSSON, Olle; DANELL, Rickard; SCHNEIDER, Wiborg Jesper. How to use Bibexcel for various types of bibliometric analysis. In: ÅSTRÖM, Fredrik; DANELL, Rickard (Ed.). **Celebrating scholarly communication studies: a Festschrift for Olle Persson at his 60th Birthday**. Leuven: Bélgica International Society for Scientometrics and Informetrics, 2009. p. 9-24.

PRITCHARD, Alan. Statistical bibliography or bibliometrics? **Journal of Documentation**, Londres, v. 25, n. 4, p. 348-349, 1969.

RAGHUPATHI, Wullianallur; NERUR, Sridhar. The intellectual structure of health and medical informatics. **International Journal of Healthcare Information Systems and Informatics**, v. 5, p. 20-34, 2010. DOI: <https://www.doi.org/10.4018/jhisi.2010100102>. Disponível em: <https://www.igi-global.com/gateway/article/47429>. Acesso em: 10 ago. 2021.

SCHUEMIE, Martijn Janssen *et al.* Mapping the domain of medical informatics. **Methods of Information in Medicine**, Stuttgart, v. 48, n. 1, p. 76-83, 2009. DOI: <https://doi.org/10.3414/me0576>. Disponível em: <https://www.thieme-connect.de/products/ejournals/abstract/10.3414/ME0576>. Acesso em: 10 ago. 2021.

SERENKO, Alexander; DOHAN, Michael S.; TAN, Joseph. Global ranking of management- and clinical-centered e-health journals. **Communications of the Association for Information Systems**, v. 41, p. 198-215, 2017. DOI: <https://doi.org/10.17705/1CAIS.04109>. Disponível em: <https://aisel.aisnet.org/cais/vol41/iss1/9/>. Acesso em: 10 ago. 2021.

SPRECKELSEN, Cord; DESERNO, Thomas; SPITZER, Klaus. Visibility of medical informatics – regarding bibliometric indices and databases. **BMC Medical Informatics and Decision Making**, [s. l.], v. 11, 2011. DOI: <https://doi.org/10.1186/1472-6947-11-24>. Disponível em: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-11-24>. Acesso em: 10 ago. 2021.

SWEILEH, Waleed M. *et al.* Bibliometric analysis of worldwide scientific literature in mobile-health: 2006-2016. **BMC Medical Informatics and Decision Making**, Londres, v. 17, p. 1-12, 2017. DOI: <https://dx.doi.org/10.1186%2Fs12911-017-0476-7>. Disponível em: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-017-0476-7>. Acesso em: 10 ago. 2021.

TOPOL, Eric Jeffrey. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, [s. l.], v. 25, p. 44-56, 2019. DOI: <https://doi.org/10.1038/s41591-018-0300-7>. Disponível em: <https://www.nature.com/articles/s41591-018-0300-7>. Acesso em: 10 ago. 2021.

VAN BEMMEL, Jan H. The structure of medical informatics: bibliography on educational courses at the Free University. **Medical Informatics**, [s. l.], v. 9, n. 3, p. 175-180, 1984. DOI: <https://doi.org/10.3109/14639238409015187>. Disponível em: <https://www.tandfonline.com/doi/abs/10.3109/14639238409015187>. Acesso em: 10 ago. 2021.

VAN ECK, Nees Jan; WALTMAN, Ludo. VOS: a new method for visualizing similarities between objects. *In*: LENZ, Hans-J.; DECKER, Reinhold (Ed.). **Advances in data analysis**. Berlin: Springer, 2007. p. 299-306. DOI: https://doi.org/10.1007/978-3-540-70981-7_34. Disponível em: https://link.springer.com/chapter/10.1007/978-3-540-70981-7_34. Acesso em: 10 ago. 2021.

VOSNER, Helena Blazun; ZELEZNIK, Danica; KOKOL, Peter. Bibliometric analysis of the International Medical Informatics Association official journals. **Informatics for Health and Social Care**, Londres, v. 44, n. 4, p. 405-421, 2019. DOI: <https://doi.org/10.1080/17538157.2018.1525734>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/17538157.2018.1525734?journalCode=imif20>. Acesso em: 10 ago. 2021.

ZIPF, George Kingsley. **Human Behavior and the Principle of Least Effort**. Cambridge: Addison-Wesley, 1949.

ZUPIC, Ivan; CATER, Tomaz. Bibliometric methods in management and organization. **Organizational Research Methods**, [s. l.], v. 18, n. 3, p. 429-472, 2015. DOI: <https://doi.org/10.1177%2F1094428114562629>. Disponível em: <https://journals.sagepub.com/doi/10.1177/1094428114562629>. Acesso em: 10 ago. 2021.