

APLICAÇÃO DE FERRAMENTAS DE MINERAÇÃO DE TEXTOS NA DESCOBERTA DE CONHECIMENTO NA ANÁLISE DE ARTIGOS CIENTÍFICOS SOBRE BRAQUITERAPIA GINECOLÓGICA

Taís Marques Cerentini¹; Patrícia da Rosa Viana²; Fabrício Edler Macagnan³; Aline Aver Vanin⁴;
Sílvio César Cazella⁵

¹ Fisioterapeuta, mestranda no Programa de Pós-Graduação em Ciências da Reabilitação pela Universidade Federal de Ciências da Saúde de Porto Alegre – UFCSPA.

² Professora no departamento de Fisioterapia, UFCSPA, Porto Alegre, RS, Brasil.

³ Professor no departamento de Fisioterapia, Programa de Ciências da Reabilitação e REMIS com ênfase em Onco-Hematologia, UFCSPA, Porto Alegre, RS, Brasil.

⁴ Professora no departamento de Educação e Humanidades, UFCSPA, Porto Alegre, RS, Brasil.

⁵ Professor no departamento de Ciências Exatas e Sociais Aplicadas, Programa de Ciência da Reabilitação, UFCSPA, Porto Alegre, RS, Brasil.

RESUMO:

Contextualização: A braquiterapia ginecológica é uma forma de radiação intracavitária com alta taxa de dose, porém pode causar efeitos secundários em tecidos adjacentes devido à radiotoxicidade. O método de mineração de texto se utiliza de ferramentas para descoberta de conhecimento em textos semi- ou não-estruturados. **Objetivo:** analisar um corpus de artigos científicos sobre braquiterapia ginecológica a partir do viés da descoberta de conhecimento em textos através de ferramentas de mineração de texto. **Métodos:** Foram aplicadas ferramentas de mineração de textos em um conjunto de treze artigos científicos. **Resultados:** Encontramos associação entre as palavras mais frequentes utilizadas no corpus com os dados na literatura sobre o assunto em questão. **Conclusão:** A utilização de ferramentas de mineração de textos semi- ou não-estruturados é eficaz na descoberta de conhecimento em braquiterapia ginecológica.

Palavras-chave: Mineração de Textos; Descoberta de Conhecimento; Braquiterapia.

INTRODUÇÃO

O câncer é um problema de saúde pública, característico de países em desenvolvimento (WILD, 2014), com comportamento neoplásico, incidência e distribuição geográfica dependente de fatores genéticos, físicos e ambientais (INCA,

2008). O câncer ginecológico inclui o câncer de cérvix uterino, de corpo uterino, de ovário, de vulva e de vagina (FAUBION et al., 2015).

Braquiterapia é uma forma de radioterapia que tem seu nome de origem grega, “Brachys” que significa “curta distância”, que consiste na colocação da fonte radioativa diretamente no tumor ou em uma área muito próxima a ele (LANCASTER, 2004; ZOMKOWSK et al., 2016). Pode ser aplicada com diferentes quantidades de material radioativo, ou seja, baixa taxa de dose (de 0,4 a 2 Greys/hora), média taxa de dose (2 a 12 Greys/hora) ou alta taxa de dose (dose superior a 0,2 Greys/minuto ou superior a 12Grys/hora) (ESTEVES et al., 2004). Na braquiterapia ginecológica, utiliza-se alta taxa de dose, conhecida pela sigla BATD, e normalmente utiliza-se a técnica intracavitária, ou seja, são colocados aplicadores variados com um ou mais canais (vaginal, intrauterino, retais, etc.). (ESTEVES et al., 2004); FERNANDES; JUNIOR, 2016).

A radiação impede a reprodução do ácido desoxirribonucleico (DNA) celular, afetando tanto células tumorais quanto células saudáveis (FRIGATO; HOGA, 2003) causando um efeito biológico da radiação no tecido que irá depender de fatores como a dose utilizada, a forma de administração, o fracionamento da dose, o tecido em que é aplicada e como este tecido absorve a radiação, entre outros(ESTEVES et al., 2004; ROSA et. al., 2016)..

Alguns tecidos tratados com BATD ou próximos ao local de tratamento, como é o caso da mucosa vaginal, podem apresentar alterações como a estenose vaginal (estreitamento e/ou encurtamento do canal vaginal), telangiectasias (formação de novos pequenos vasos), dispareunia (dor na relação sexual) e secura vaginal. (THOMAS, 1999; MOHAMED et al., 2016). Outros fatores, como irritabilidade vesical, diarreia, alterações cutâneas, fístulas intestinais e vesicais e fibrose vaginal, também podem ser inevitáveis ao tratamento (ROSA et al., 2016).

Devido a grande quantidade de textos sobre BATD, uma alternativa para buscar resultados que evidenciem as relações entre os termos é com o uso de ferramentas de mineração texto. A análise de dados em formato semi ou não estruturado pode ser mais complexa e demandar mais tempo quando comparada a dados estruturados, exatamente por possuírem essa característica de não estruturação. Consequentemente, técnicas e ferramentas para tratamento deste tipo de dado facilitaria a descoberta de conhecimento e otimizaria o tempo do leitor (MORAIS; AMBRÓSIO, 2007).

Text Mining (TM) é um método para descoberta de conhecimento que utiliza etapas de processamento de linguagem natural, gerenciamento de conhecimento, mineração de dados e aprendizado de máquina em texto livre, ou seja, dados semi ou não estruturados (LUCINI et al., 2017), e vem sendo amplamente utilizado para representar informações sobre o câncer (SPASIĆ et al., 2014).

Consiste de quatro etapas: coleta de documentos, também conhecida como formação do corpus, pré-processamento dos textos, extração de conhecimento e pós-processamento. A primeira etapa é composta pela seleção de dados que irá compor a base de textos de interesse e a necessidade de um especialista no domínio da pesquisa que selecione os artigos que melhor expressam o conteúdo, buscando identificar similaridades em função da morfologia e do significado dos termos nos textos. Esse conjunto de texto irá compor o que chamamos de *corpus*.

Após a construção do *corpus*, inicia-se a etapa de pré-processamento, na qual o corpus é transformado em uma representação que poderá ser utilizada pelos algoritmos de extração de padrões, normalmente uma tabela atributo-valor que será utilizada na fase de extração de padrões, que, depois de extraídos, serão analisados na fase de pós-processamento. Ainda no pré-processamento, palavras que são consideradas irrelevantes no texto, ou *stopwords*, são eliminadas. Normalmente, são consideradas *stopwords* preposições, pronomes, artigos, advérbios e outras classes de palavras auxiliares (ARANHA; PASSOS, 2006; MORAIS; AMBRÓSIO, 2007).

No pós-processamento, os padrões obtidos na fase de extração serão avaliados e validados, em que normalmente se usam medidas objetivas para avaliar o modelo obtido, sendo ainda de suma importância a avaliação do especialista do domínio para validar o modelo (HONORATO, 2008). Nesse sentido, será possível extrair conhecimento útil para o usuário final (MORAIS; AMBRÓSIO, 2007).

Diante disso, este estudo tem como objetivo analisar artigos científicos sobre braquiterapia ginecológica na descoberta de conhecimento em textos através de ferramentas de mineração de texto.

MATERIAIS E MÉTODOS

Neste estudo, aplicamos os métodos de mineração de textos para analisar um *corpus* composto por treze artigos científicos, buscando descobrir conhecimento sobre braquiterapia ginecológica.

Primeiramente, foi necessário eleger os artigos científicos que viriam compor o *corpus*. Para tanto, utilizamos a base de artigos científicos PubMed¹, na qual realizamos a pesquisa com as seguintes palavras-chave: *Vaginal stenosis*; *vaginal constriction*; e *brachytherapy*. Com base nessas palavras-chave foram elaboradas as seguintes *strings* de busca: “*Vaginal stenosis OR vaginal constriction AND brachytherapy*”, em que foi estipulado que estas expressões poderiam aparecer em quaisquer campos indexados (por exemplo, no título, no resumo, nas palavras-chave, etc.). Os resultados da busca foram analisados por um especialista do domínio e foram selecionados, através da leitura dos resumos dos artigos retornados, os treze primeiros artigos científicos com maior relação com a pesquisa conduzida, visando compor o *corpus*.

Iniciamos a fase de pré-processamento com a transformação dos arquivos em formato PDF para o formato TXT através da ferramenta AntConverter. Após a transformação, removemos as *stopwords*, e foram retirados o cabeçalho inicial, dados dos autores, resumo e referências. Para dar continuidade ao pré-processamento dos textos, utilizamos as ferramentas listadas na Tabela 1.

Tabela 1 Ferramentas utilizadas na mineração dos textos.

Nome da ferramenta	Descrição da ferramenta	Utilização da ferramenta na pesquisa
<i>AntConverter</i> ²	Disponível para download gratuito converter arquivos PDF e Word (DOCX) em texto simples para uso em ferramentas de processamento do <i>Corpus</i> .	Utilizamos para converter os treze artigos PDF em arquivo TXT para realização do pré-processamento
<i>WordCounter</i> ³	Online e gratuito, ele calcula o número de palavras, frases e caracteres, entre outros aspectos.	Analizamos a incidência de termos de pesquisa e termos mais gerais (Tabela 3).
<i>TagCrowd</i> ⁴	Online e gratuito, permite criar nuvens de marcadores (tags nuvens).	Utilizamos para visualizar a nuvem de palavras mais frequente no corpus (Figura 3).

¹ <https://www.ncbi.nlm.nih.gov/pubmed/>

² <http://www.laurenceanthony.net/software.html>

³ <https://wordcounter.net/>

⁴ <http://tagcrowd.com/>

Voyant Tools ⁵	Online e gratuito, analisa o texto e cria diversos formatos de gráficos e análises.	Utilizamos a ferramenta <i>summary</i> para visualizar as palavras mais frequentes em cada artigo (tabela 2) e também a contagem do número total de palavras (tabela 4); A ferramenta <i>reader</i> que apresenta os gráficos de coluna (figura 1); A ferramenta <i>links</i> que apresenta as correlações das palavras (figura 2);
---------------------------	---	---

Ainda na fase de pré-processamento, utilizamos as ferramentas *WordCounter* para obter as dez palavras mais frequentes no *corpus*, bem como o número de vezes e a frequência com que aparecem no *corpus*. Utilizamos a ferramenta TagCrowd para formar uma nuvem que apresentasse os principais termos do texto de forma mais visual, enfatizando palavras e termos mais frequentes. Já a *Voyant Tools* é um ambiente de análise de texto bastante completo, que engloba várias ferramentas, das quais utilizamos três em nosso estudo. A primeira delas, a *Summary*, gera uma visão geral simples e textual do *corpus*, incluindo número de palavras, número de palavras únicas, documentos mais longos e mais curtos, maior e menor densidade de vocabulário, número médio de palavras por sentença, palavras mais frequentes, picos notáveis de frequência e palavras distintivas. Em nosso estudo, utilizamos o *Summary* apenas para a contagem total de palavras do *corpus*. Outra ferramenta presente no *Voyant Tools* que utilizamos em nosso estudo é a *Reader*, que fornece uma maneira de ler documentos no corpus e permite buscas no texto conforme necessário. Também gera um gráfico de colunas sobre a participação de cada documento no *corpus* que utilizamos em nosso trabalho para mostrar a importância e contribuição de cada artigo selecionado dentro do *corpus*. Com a ferramenta Collocates Graph, conseguimos enxergar as conexões existentes entre as palavras, porque seu objetivo é representar as palavras-chave e termos mais frequentes em proximidade como um gráfico de rede direcionado por força.

Os resultados da pesquisa estão resumidos na Tabela 1, que mostra detalhes como ano de publicação, país de origem, periódico ou evento em que foi publicado e uma identificação (ID) que foi atribuído a cada trabalho para facilitar a referência do mesmo no decorrer deste trabalho. Estes critérios foram escolhidos para melhor situar o leitor onde se pesquisa e em que local é publicado sobre o assunto.

⁵ <https://voyant-tools.org/>

Tabela 2 Artigos selecionados para composição do corpus.

ID	Título do artigo	Ano Publicação	País de origem	Periódico de publicação	Número de palavras
1	A nurse-led sexual rehabilitation intervention after radiotherapy for gynecological cancer	2017	Holanda	Support Care Cancer	4,314
2	Comparisons of Late Vaginal Mucosal Reactions Between Interstitial and Conventional Intracavitary Brachytherapy in Patients with Gynecological Cancer: Speculation on the Relation Between Pallor Reaction and Stenosis	2013	Japão	Anticancer Research	3,025
3	Dose–effect relationship and risk factors for vaginal stenosis after definitive radio(chemo)therapy with image-guided brachytherapy for locally advanced cervical cancer in the EMBRACE study	2016	Austrália	Radiotherapy and Oncology	3,664
4	Dose-volume correlation of cumulative vaginal doses and late toxicity after adjuvant external radiation and brachytherapy for cervical cancer	2017	India	Brachytherapy	3,102
5	Determination Of Prognostic Factors For Vaginal Mucosal Toxicity Associated With Intravaginal High-dose Rate Brachytherapy In Patients With Endometrial Cancer	2010	EUA	International Journal of Radiation Oncology	4,637
6	High dose rate 192Ir afterloading brachytherapy for cancer of the vagina	2003	EUA	The British Journal of Radiology	5,026
7	Perceptions and practices regarding women's vaginal health following radiation	2017	EUA	American Society for Radiation	3,777

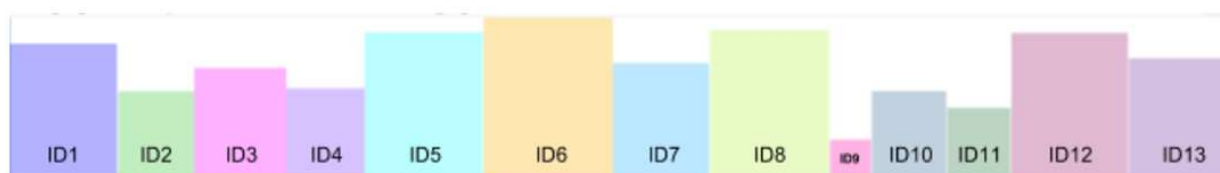
	therapy: a survey of radiation oncologists practicing in the United States			Oncology	
8	Preventing vaginal stenosis after brachytherapy for gynaecological cancer: an overview of Australian practices	2004	Austrália	European Journal of Oncology Nursing	4,688
9	Prevention of Vaginal Stenosis in Patients Following Vaginal Brachytherapy	1999	Reino Unido	Clinical Oncology	1,698
10	Radiation-induced vaginal stenosis: current perspectives	2017	Austrália	International Journal of Women's Health	3,011
11	Role of vaginal pallor reaction in predicting late vaginal stenosis after high-dose-rate brachytherapy in treatment-naïve patients with cervical cancer	2015	Japão	Journal of Gynecologic Oncology	2,574
12	Severe Late Toxicities Following Concomitant Chemoradiotherapy Compared to Radiotherapy Alone in Cervical Cancer: An Inter-era Analysis	2012	EUA	International Journal of Radiation Oncology	4,614
13	Vaginal dose de-escalation in image guided adaptive brachytherapy for locally advanced cervical cancer	2016	Dinamarca	Radiotherapy and Oncology	3,917

A Tabela 2 mostra que, dos 13 artigos selecionados, quatro são dos Estados Unidos da América, dois do Japão e dois da Austrália, o que mostra que esse assunto e temas podem ser mais discutidos nesses países em comparação com os outros países que apresentam uma publicação cada. Esta informação é importante por facilitar a procura por artigos científicos e até mesmo a tomada de decisão de uma revista para publicação, pois nos mostra onde se está falando sobre o tema e

que é um assunto relevante e/ou necessário de se buscar mais informações. Outro dado que obtemos é sobre as revistas de publicação: apenas duas aparecem na tabela duas vezes, ou seja, apenas duas revistas publicaram sobre o mesmo tema e ele é um assunto abordado em várias revistas diferentes, mostrando-se um assunto com potencial interesse tanto para pesquisa quanto para a clínica.

Por fim, na fase de pós-processamento, observamos padrões de textos gerados como saída pelas ferramentas de mineração de texto utilizadas, e analisamos a ocorrência de termos para cada um dos treze artigos científicos e depois para o *corpus*.

Figura 1 Visualização do corpus. Cada cor corresponde a uma publicação. No total, o corpus foi formado por 48,047 palavras.



Após a apresentação da metodologia, a próxima seção apresentará os resultados alcançados através destas ferramentas, bem como a análise e discussão baseada em evidências.

RESULTADOS E DISCUSSÃO

Em primeira análise, observamos as palavras e/ou termos mais frequentes no *corpus composto de 48,047 palavras*, que serão apresentados na Tabela 3.

Tabela 3 Dez Palavras/termos frequentes no corpus apresentados pela ferramenta WordCounter.

Palavra/termo	Número de vezes que aparece	Frequência que aparece
Vaginal	888	7%
Dose	413	3%

Patients	410	3%
Brachytherapy	250	2%
Stenosis	245	2%
Gy(gray)	225	2%
Treatment	210	2%
After	188	1%
Sexual	187	1%
Cancer	176	1%

Diante da análise do *corpus*, podemos perceber palavras que não eram esperadas, mas se mostraram relevantes. Conhecendo o domínio da pesquisa, podemos interpretar esses achados enxergando a conexão existente entre essas palavras, apesar de apenas duas delas terem sido nossas palavras-chave: “vaginal stenosis” e “brachytherapy”. Estudos anteriores comprovam que há uma correlação positiva entre a dose de radiação utilizada, ou seja, quanto mais grays(Gy) a mulheres receber, maior o risco de apresentar efeitos tardios secundários à radiação, como é o caso da estenose vaginal, Isso justifica a alta frequência da palavra *after* no *corpus*. Portanto, ao analisarmos esses termos, percebemos sua interdependência conforme publicações científicas que comprovam a relação (BAHNG et al., 2010; KIRCHHEINER et al., 2016; MOHAMED et al., 2017; SINGH et al., 2017).

Outro termo que aparece com frequência no *corpus* é “*sexual*”. Para LANCASTER (2004); PARK et al. (2015); KIRCHHEINER et al, (2016), mulheres submetidas a tratamento radioterápico para câncer ginecológico, começam a apresentar alguns sintomas, como estreitamento da vagina e perda da lubrificação, formação de aderências e fibroses e a perda da elasticidade vaginal, o que levará a uma disfunção sexual .

Com a ferramenta *Voyant Tools* elaboramos a Tabela 4, que apresenta quais termos aparecem em cada artigo do *corpus* e suas respectivas frequências.

Tabela 4 Palavras/termos que mais aparecem em cada ID.

ID	Palavra/termo1	Palavra/termo 2	Palavra/termo 3	Palavra/termo4	Palavra/termo 5
1	Sexual * (70)	Participants (41)	Months (40)	Brachytherapy (36)	pelvic external beam radiation (36)
2	Patients (55)	Greys ∞ (39)	Interstitial brachytherapy (35)	Vaginal ¥ (29)	conventional intracavitary brachytherapy (27)
3	Dose (101)	Vaginal ¥ (88)	Point (36)	Stenosis \pm (36)	Risk (32)
4	Vaginal ¥ (136)	Dose (59)'	Patients (36)	Sexual * (32)	Brachytherapy (27)
5	Vaginal ¥ (91)	Toxicity (52)	Dilator (46)	Patients (49)	Use (40)
6	Patients (67)	Dose (54)	Greys ∞ (48)	Vaginal ¥ (44)	Brachytherapy (41)
7	Vaginal ¥ (71)	Vaginal Stenosis \approx (65)	use (35)	Survey (32)	Respondents (31)
8	Vaginal ¥ (113)	-	-	Stenosis \pm (68)	Use (64)
9	Stent (39)	Vaginal ¥ (38)	Patients (33)	Stenosis \pm (27)	Use (18)
10	Vaginal Stenosis \approx (61)	Radiotherapy (36)	Vaginal ¥ (59)	Cancer (19)	Grade Σ (18)
11	Stenosis \pm (46)	Grade Σ (45)	Vaginal ¥ (44)	Pallor (29)	Reaction (24)
12	Severe late toxicities (85)	Vaginal ¥ (66)	-	Grade Σ (46)	Patients (44)
13	Vaginal ¥	Dose	Vaginal dose	Greys ∞	Loading

	(113)	(101)	de-escalation (70)	(43)	(33)
--	-------	-------	-----------------------	------	------

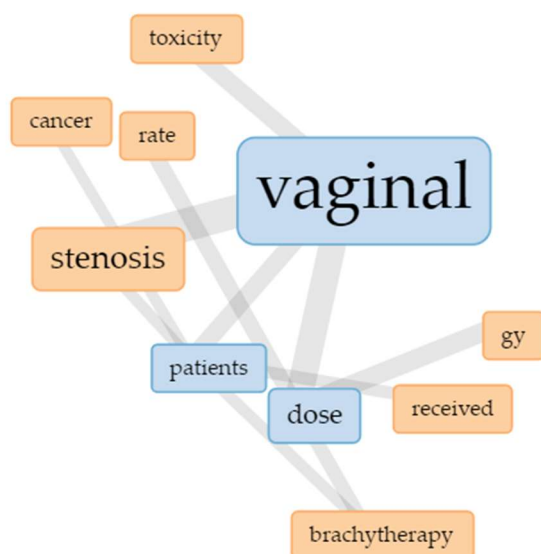
Uma observação importante na Tabela 4 é que o símbolo representado pelo menos (-) significa stopwords que permaneceram no texto e que, em análise por um especialista do domínio, foram consideradas irrelevantes dentro do contexto. Também separamos por cores as mesmas palavras para sabermos a frequência de cada uma em cada artigo e no *corpus*. Podemos confirmar nossos achados anteriores, sobre a influência da dose de radiação utilizada e sua relação dose/efeito com o aparecimento da estenose vaginal, indo ao encontro de Singh et al. (2017). A tabela acima ainda demonstra que corroboramos outro achado importante: a vagina sofre de efeitos tardios a essa radiotoxicidade.

A ocorrência de termos como “pallor”, “toxicity” e “sexual” direcionam o olhar do profissional de saúde para sinais e sintomas que a mulher pode apresentar em decorrência da radiação pélvica. Segundo Yoshida et al. (2015), a presença de palidez vaginal foi fator preditivo para um maior risco de a mulher apresentar estenose vaginal como efeito tardio, bem como a alta taxa de dose de radiação.

As medidas para avaliar a estenose vaginal são bastante controversas, não havendo um consenso entre autores nem padronização de um método, dificultando a medida de incidência real da estenose e prevenção da mesma (SILVA et al., 2010; ROSA et al., 2016). Porém, através desta ferramenta de mineração de textos, podemos perceber que nos ID 10, 11, 12 uma frequente aparição da palavra *grade*, que é um modelo de escala para a graduação de estenose vaginal proposto por Brand et al. (2006).

A Figura 2 é uma representação gráfica dos termos mais relevantes do corpus, usando a ferramenta Collocates Graph presente no *Voyant tools*.

Figura 2 Representação gráfica dos termos mais relevantes no corpus.



É possível observar as conexões entre os termos com maior frequência no corpus e associar os termos “vaginal” com “stenosis”, “stenosis” com “brachytherapy”, o que mostra que a maior incidência de estenose em mulheres que realizam braquiterapia, indo ao encontro do estudo de Brand et al. (2006), que encontraram uma frequência de 38% de estenose vaginal em mulheres que realizaram braquiterapia. A conexão entre as palavras também nos mostra a fidedignidade da ferramenta, pois sabemos que a estenose vaginal (*vaginal; stenosis*) é uma toxicidade (*toxicity*) que ocorre em pacientes (*patients*) que receberam (*received*) uma alta taxa (*rate*) de dose (*dose/gy*) por braquiterapia (*brachytherapy*) para o tratamento do câncer (*cancer*) ginecológico. Ou seja, podemos ter uma visão geral do *corpus* apenas analisando as conexões propiciadas pela ferramenta Collocates Graph, o que nos mostra que as ferramentas de mineração de textos podem otimizar a leitura.

Em conclusão, a Figura 3 mostra a nuvem de palavras gerada pelo TagCrowd com os termos mais frequentes no corpus, possibilitando visualizar instantaneamente os assuntos abordados e possibilitando a obtenção de um panorama do texto como um todo.

Figura 3 Nuvem com palavras mais frequentes no Corpus.



Nesse sentido, foi possível observar que, através da automatização da análise de um grande volume de texto, é possível entender a ideia geral e suas correlações de uma forma mais rápida e eficiente do que seria manualmente, através da leitura de cada documento, otimizando o tempo do leitor. No caso deste estudo, verificamos a importância do uso dessas ferramentas na análise do texto científico voltado para a clínica. Atualmente, a braquiterapia ginecológica é um tratamento difundido por todo o mundo no combate ao câncer devido à sua precisão e à alta tecnologia (ESTEVEZ ET AL., 2004). Porém, devido à radiobiologia de alguns tecidos, sabe-se que podem ocorrer danos secundários, evidenciados neste trabalho através da alta frequência de palavras como *toxicity*, presente na análise da Figura 3.

Outras palavras que são apresentadas com alta frequência ao longo deste trabalho são palidez vaginal, estenose vaginal, dentre outros danos à mucosa. Assim, concluímos que esta automatização na análise do texto auxiliará profissionais da área a direcionar um olhar para essas comorbidades, servindo de alerta na prevenção destas complicações e também no direcionamento de novas pesquisas sobre o assunto em questão.

CONSIDERAÇÕES FINAIS

Neste trabalho demonstramos a possibilidade de usar técnicas e ferramentas de mineração de textos na descoberta de conhecimento à comunidade acadêmica e

na prática clínica do profissional de saúde em *corpus* de textos semi- ou não estruturados com ferramentas de mineração de textos.

Os trabalhos selecionados abordaram aspectos como o uso do dilatador, função sexual, a dose em greys de radiação e a toxicidade apresentada pela mucosa, termos que se relacionam entre si e com as palavras-chave, descobrindo assim tendências nas mulheres que realizam este tipo de tratamento.

Pode-se concluir que as ferramentas de mineração de textos compõem um sistema simples, rápido, automatizado e eficiente de analisar um volume de dados considerável, otimizando o tempo para organização desses dados...

REFERÊNCIAS:

ANT CONVERTER. Disponível em: <http://www.laurenceanthony.net/software.html>. Acesso em: 11.06.2017.

ARANHA, Christian; PASSOS, Emmanuel. A tecnologia de mineração de textos. Revista Eletrônica de Sistemas de Informação ISSN 1677-3071 doi: 10.21529/RESI, v. 5, n. 2, 2006.

BAHNG, Agnes Y. et al. Determination of prognostic factors for vaginal mucosal toxicity associated with intravaginal high-dose rate brachytherapy in patients with endometrial cancer. International Journal of Radiation Oncology* Biology* Physics, v. 82, n. 2, p. 667-673, 2012. BRAND, Alison H.; DO, Viet; STENLAKE, Annie. Can an educational intervention improve compliance with vaginal dilator use in patients treated with radiation for a gynecological malignancy?. International Journal of Gynecological Cancer, v. 22, n. 5, p. 897-904, 2012.

CULLEN, Kimberley et al. Toward clinical care guidelines for supporting rehabilitative vaginal dilator use with women recovering from cervical cancer. Supportive Care in Cancer, v. 21, n. 7, p. 1911-1917, 2013.

ESTEVES, Sérgio Carlos Barros et al. Braquiterapia de alta taxa de dose no Brasil. Radiol Bras, v. 37, n. 5, p. 337-41, 2004.

FERNANDES, Jéssica Alves; JUNIOR, Paulo Pinhal. TÉCNICAS UTILIZADAS EM RADIOTERAPIA. UNILUS Ensino e Pesquisa, v. 13, n. 30, p. 246, 2016.

HONORATO, Daniel de Faveri. Metodologia para mapeamento de informações não estruturadas descritas em laudos médicos para uma representação atributo-valor. 2008. Tese de Doutorado. Universidade de São Paulo.

KIRCHHEINER, Kathrin et al. Dose–effect relationship and risk factors for vaginal

stenosis after definitive radio (chemo) therapy with image-guided brachytherapy for locally advanced cervical cancer in the EMBRACE study. *Radiotherapy and Oncology*, v. 118, n. 1, p. 160-166, 2016.

LANCASTER, Letitia. Preventing vaginal stenosis after brachytherapy for gynaecological cancer: an overview of Australian practices. *European Journal of Oncology Nursing*, v. 8, n. 1, p. 30-39, 2004.

LUCINI, Filipe R. et al. Text mining approach to predict hospital admissions using early medical records from the emergency department. *International Journal of Medical Informatics*, 2017.

MILES, Tracie; JOHNSON, Nick. Vaginal dilator therapy for women receiving pelvic radiotherapy. *The Cochrane Library*, 2014.

MOHAMED, Sandy et al. Vaginal dose de-escalation in image guided adaptive brachytherapy for locally advanced cervical cancer. *Radiotherapy and Oncology*, 2016.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. Mineração de textos. *Relatório Técnico—Instituto de Informática (UFG)*, 2007.

PARK, Henry S. et al. Predictors of vaginal stenosis after intravaginal high-dose-rate brachytherapy for endometrial carcinoma. *Brachytherapy*, v. 14, n. 4, p. 464-470, 2015.

ROSA, Luciana Martins da et al. Evaluation and classification of vaginal stenosis after brachytherapy. *Texto & Contexto-Enfermagem*, v. 25, n. 2, 2016.

SILVA, M.P.P et al., Métodos de Evaluación; POSRADIOTERAPIA, Vaginal. Métodos Avaliativos para Estenose Vaginal Pós-Radioterapia. *Revista Brasileira de Cancerologia*, v. 56, n. 1, p. 71-83, 2010.

SINGH, Roshni et al. Dose-volume correlation of cumulative vaginal doses and late toxicity after adjuvant external radiation and brachytherapy for cervical cancer. *Brachytherapy*, 2017.

SPASIĆ, Irena et al. Text mining of cancer-related information: review of current status and future directions. *International journal of medical informatics*, v. 83, n. 9, p. 605-623, 2014.

VOYANT TOOLS. Disponível em: <https://voyant-tools.org/>. Acesso em: 12.06.2017.

TAG CROWD. Disponível em: <http://tagcrowd.com/>. Acesso em: 12.06.2017.

WORD COUNTER. Disponível em: <https://wordcounter.net/>. Acesso em 12.06.2017.

YOSHIDA, Ken et al. Role of vaginal pallor reaction in predicting late vaginal

stenosis after high-dose-rate brachytherapy in treatment-naive patients with cervical cancer. *Journal of gynecologic oncology*, v. 26, n. 3, p. 179-184, 2015.

.