XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB

ISSN 2177-3688

GT-8 – Informação e Tecnologia

MAPEAMENTO DE PRODUTOS UTILIZANDO MINERAÇÃO DE TEXTOS E NUVEM DE PALAVRAS

PRODUCT MAPPING USING TEXT MINING AND WORD CLOUD

Daniela Souza Moreira da Silva. UFSC.

Eduardo Lima Leite Nascimento. UFSC.

Angel Freddy Godoy Vieira. UFSC.

Dennis Kerr Coelho. UNIVALI.

Modalidade: Trabalho Completo

Resumo: Visando identificar quais são os produtos mais ofertados no setor de capacitação empresarial, bem como, aqueles que têm maior peso no mercado e não são oferecidos por uma empresa de capacitação, foi realizado um estudo em uma base de dados de cursos construída para analisar os dados utilizando mineração de textos. Primeiramente foi necessário definir as áreas de interesse da pesquisa: Financeira, Mercado, Pessoas, Produção, Estratégia e Legal/Jurídica, somente cursos profissionalizantes. Depois, foram mapeados os concorrentes e os cursos. Os dados coletados foram o nome do curso, cidade, estado, área, tipo de oferta, subárea, faixa de preço, duração e maturidade. Com os dados coletados foi gerado o corpus textual da pesquisa, realizada a normalização dos dados por meio da padronização dos caracteres em maiúsculo e removidas as stopwords. Em seguida, foi calculada a métrica de similaridade dos termos com o cálculo de distância de Levenshtein e foram contabilizados os termos mais frequentes e os produtos (títulos) mais semelhantes, apresentando o resultado em uma nuvem de palavras e em uma nuvem de produtos semelhantes para cada área. Por meio do tratamento textual realizada no corpus construído para pesquisa e da apresentação dos resultados com as nuvens de palavras foi possível apresentar à empresa de capacitação empresarial quais são os produtos que ela e seus concorrentes estão oferecendo em maior volume, bem como, quais produtos somente a concorrência oferta e tem indícios para possíveis investimentos e revisão do seu portfólio.

Palavras-Chave: Nuvem de Palavras. Mineração de Textos. Mapeamento de Concorrentes.

Abstract: In order to identify which are the most offered products in the business training sector, as well as those that have greater weight in the market and are not offered by a training company, a study was carried out in a database of courses built to analyze the data using text mining. First, it was necessary to define the areas of interest for the research: Finance, Market, People, Production, Strategy and Legal/Legal, only professional courses. Afterwards, the competitors and the courses were mapped. The data collected were the course name, city, state, area, type of offer, subarea, price range, duration and maturity. With the collected data, the textual corpus of the research was generated, the data normalization was carried out through the standardization of the characters in capital letters and

the stopwords were removed. Then, the similarity metric of the terms was calculated with the Levenshtein distance calculation and the most frequent terms and the most similar products (titles) were counted, presenting the result in a cloud of words and in a cloud of similar products for each area. Through the textual treatment carried out in the corpus built for research and the presentation of the results with the clouds of words, it was possible to present to the business training company which products it and its competitors are offering in greater volume, as well as which products only the competition offers and has indications for possible investments and review of its portfolio.

Keywords: Word Cloud. Text Mining. Competitor Mapping.

1 INTRODUÇÃO

Considerada uma evolução da área de RI (recuperação de informações) a mineração de textos é um processo para descobrir conhecimento por meio de técnicas de análise e extração a partir de textos ou palavras. São utilizados algoritmos que processam os textos e identificam informações úteis ainda que estejam armazenados de forma não estruturada (MORAIS; AMBRÓSIO, 2007).

A mineração de textos é diferente do mecanismo de busca realizado pelo usuário, uma vez que, ele sabe o que deseja encontrar, enquanto na mineração de textos ela busca por informações desconhecidas (ARANHA; PASSOS, 2006).

A tecnologia de mineração de textos possui grande importância e contribui pelo fato de o grande volume de informações estarem armazenados no formato de texto. Segundo Barion e Lago (2008) estima-se que mais de 80% das informações estão armazenadas como texto.

Para identificar as principais tendências no setor de capacitação empresarial de uma empresa específica em relação a seus concorrentes surgiu a necessidade de mapear quais são os produtos (cursos, treinamentos, consultorias) que estão sendo ofertados no mercado e quais possuem maior peso.

Foi gerada uma base de dados textual a partir do monitoramento do mercado realizado por meio de consultas em diversos sites da internet, buscando informações de seis áreas de interesse para pesquisa, são elas: financeira, mercado, pessoas, produção, estratégia e legal/jurídica.

Ao realizar a coleta dos dados deparou-se com um problema relacionado ao nome dos produtos oferecidos, pois eles possuem nomes semelhantes e não tem um padrão específico. Diante deste contexto, foi definida a utilização da mineração de textos para extração de informações através dos nomes de produtos.

O objetivo geral deste estudo foi apresentar, por meio de nuvens de palavras, os cursos que são ofertados por uma empresa de capacitação e seus concorrentes, bem como, os cursos que são oferecidos somente pelos seus concorrentes. Como objetivos específicos este estudo propõe: i) Construir a base textual a partir de dados coletados, ii) implementar o algoritmo para mineração dos produtos e iii) implementar a metodologia para o cálculo de similaridade de produtos para possibilitar a retirada de produtos semelhantes aos produtos da empresa e que são oferecidos pelos concorrentes e os produtos que não ofertados pela empresa e são oferecidos pelos concorrentes.

2 DESENVOLVIMENTO

Nesta seção será apresentado o referencial teórico do trabalho, seguido da metodologia definida para a pesquisa. Em seguida serão apresentados os resultados obtidos e as análises realizadas.

2.1 Fundamentação Teórica

A mineração de textos é uma extensão da mineração de dados voltada para dados textuais, cujo objetivo é extrair informações significativas a partir deste tipo de dado: texto livre ou semiestruturado (IGNOATTO; WEBBER, 2019). Isto é, o processo da descoberta de conhecimento por meio da extração automatizada de informações em dados textuais não estruturados (NOTA; POSTIGIOLNE; CARVELLO, 2022).

Há um grande volume de dados sendo gerados diariamente e eles não são, necessariamente, armazenados em bancos de dados relacionais. Conforme citado por Gonçalves (2012) há diversas informações disponibilizadas digitalmente no formato de texto, tais como: jornais, revistas, páginas da web, redes sociais, blogs, e-mails, arquivos pdf, arquivos XML (*eXtensible Markup Language*), arquivos HTML (*HyperText Markup Language*), arquivos JSON (*JavaScript Object Notation*).

Para os dados estruturados e armazenados em bancos relacionais forma de recuperação e manipulação dos dados é por meio do SQL (*Structured Query Language*), que é uma linguagem de consulta poderosa que possui uma forma definida para acessar os dados (BARBOZA; FREITAS, 2018)

No entanto, quando os dados são textuais a forma para descobrir informações não ocorre da mesma maneira que é realizada na exploração de dados estruturados. Os textos

podem estar em um formato livre, isto é, em linguagem natural o que implica transferir para o computador a tarefa de interpretar os dados realizando uma análise sintática e semântica do texto. A análise automática de textos é chamada de processamento de linguagem Natural (PNL) que envolve a execução de um algoritmo visando extrair informações (CHIARELLO *et al*, 2021).

Para os textos semi estruturados, como são os arquivos XML, HTML, JSON, eles possuem *tags* que auxiliam os algoritmos a identificar o texto de acordo com a sua localização. Ainda assim, a complexidade para descobrir o conhecimento em cima destes dados é alta, sendo necessário preparar os dados e aplicar os algoritmos de acordo com o problema a ser tratado.

A mineração de texto também chamada de análise de texto, segundo Kumar, Kar e Ilavarasan (2021), é uma técnica de inteligência artificial que converte dados não estruturados em dados estruturados utilizando PLN para aprimorar a análise utilizando algoritmos de aprendizado de máquina.

As etapas da mineração de texto são basicamente: Pré-processamento textual, processamento textual e pós- processamento textual.

Na etapa de pré-processamento, também denominada de processamento de linguagem natural, é realizada a coleta dos dados textuais e a normalização do texto.

As etapas mais comuns de normalização de texto são: limpeza textual, uniformização maiúsculas e minúsculas, remoção de símbolos e pontuação, tokenização do texto, expansões e contrações, remoção de *stopwords*, correção de grafia e lematização/radicalização.

Para mostrar o resultado dos pós processamento pode-se utilizar a representação de nuvens de palavras para visualizar os termos com maior relevância dentro de um determinado corpus textual.

No contexto organizacional, a mineração de texto pode ser empregada para diversas finalidades. Cordella *et al* (2020) investigaram a cultura de formação de professores e alunos em uma universidade italiana para compreender o processo de desenvolvimento de habilidades profissionais e o acesso ao mercado de trabalho, e ressaltam a eficácia de técnicas de mineração de texto para explorar a cultura da formação universitária.

Galati e Bigliardi (2019) abordam oportunidades e implicações em termos de educação, treinamento e trabalho na contemporaneidade, para identificar lacunas e incompatibilidades de competências, utilizando técnicas de mineração de texto.

Estudo que busca analisar as capacidades profissionais exigidas com base em anúncios de emprego usando a técnica de mineração de texto (CHUNG; CHEN, 2019).

Leon *et al* (2018) analisaram conteúdo de anúncios de emprego usando mineração de texto para identificar habilidades e competências necessárias para empregos na área de negócios.

Na próxima seção será apresentada a metodologia utilizada para realizar o mapeamento de cursos oferecidos por instituições educacionais, que sejam similares e/ou, indiquem alguma oportunidade de oferta de cursos.

2.2 Metodologia

A caracterização dessa pesquisa quanto a sua natureza é aplicada, pois busca gerar conhecimento a partir da análise de dados de uma base textual e fornecer informações para apoiar a tomada de decisão. Quanto a abordagem trata-se de uma pesquisa quantitativa que coleta e analisa os dados quantitativos. Em relação aos objetivos é um trabalho exploratório que utiliza como procedimento o estudo de caso de uma empresa de capacitação empresarial para analisar a utilização da técnica de mineração de textos e nuvem de palavras no mapeamento de produtos ofertados por empresas concorrentes.

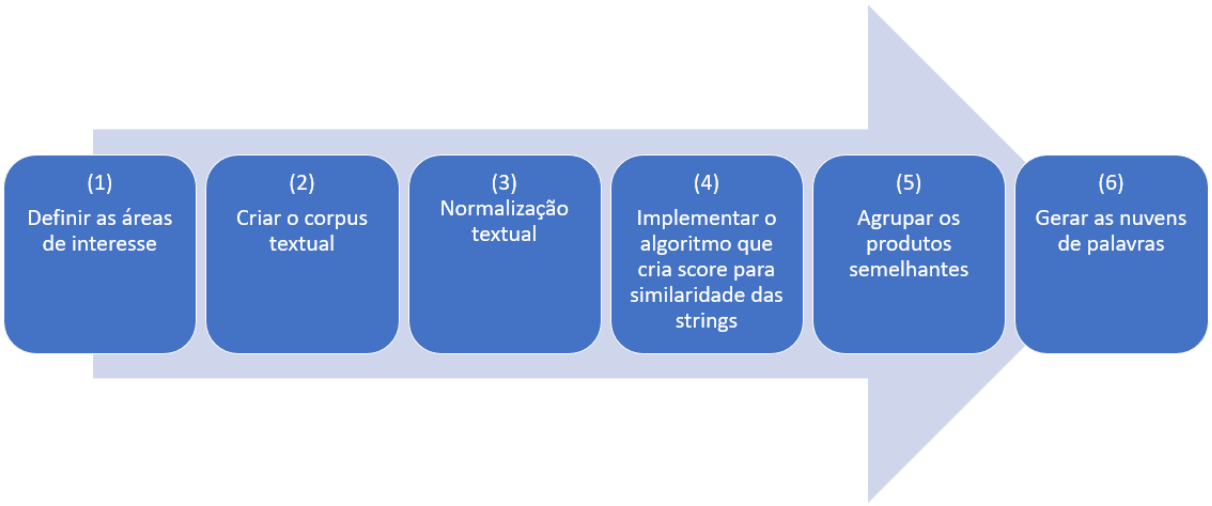
No contexto desta pesquisa ficou definido que um produto pode ser: curso, consultoria ou capacitação. Para buscar as informações dos concorrentes da empresa de capacitação empresarial foi realizada uma pesquisa na internet por concorrentes diretos que ofereciam produtos semelhantes. A extensão da pesquisa compreendeu todo o território nacional e na coleta dos dados foram consideradas as modalidades de cursos online, presenciais e semipresenciais.

A Figura 1 apresenta, de forma resumida, as atividades realizadas para gerar as nuvens de palavras que representam os produtos que são mais oferecidos no mercado pelo cliente e seus concorrentes.

Foram definidas de seis (6) grandes áreas de interesse, são elas: financeira, mercado, pessoas, produção, estratégia e legal/jurídica e cursos com foco na empresa ou no

empreendedor. Cursos técnicos voltados para o crescimento profissional não foram considerados. Essa etapa corresponde ao passo (1) da Figura 1.

Figura 1. Fluxo de atividades para gerar nuvem de palavras de cursos



Fonte: Elaborado pelo(a) autor(a)

Em seguida, foi iniciada a etapa de pré-processamento textual que engloba a coleta de dados e normalização do corpus. A construção da base (corpus textual) foi realizada por meio da coleta manual de dados na internet, no período de novembro de 2021 a janeiro de 2022. Para cada curso foram coletadas as seguintes informações: nome, cidade, estado, área, tipo de oferta, subárea, faixa de preço, duração e maturidade. Sendo o nome o atributo textual principal onde a análise e extração de informação será realizada. Essa etapa corresponde ao passo (2) Figura 1.

Para realizar a normalização do corpus, passo (3) da Figura 1, foi implementado um algoritmo em java que transforma todos os caracteres em maiúsculos e remove as palavras classificadas como *stopwords* da base de dados.

O algoritmo implementado cria um score relacionando a diferença das *strings*, e leva em consideração a existência de uma *string* dentro da outra e a existência de palavras que não modificam a identificação do curso, como por exemplo: gestão de pessoas e gestão de pessoas 2.0. Neste caso, o curso será considerado o mesmo, ainda que tenha alguma variação no nome.

Para criar os grupos de textos e analisar as *strings* foi necessário realizar a comparação das *strings* (nomes dos cursos) quanto à similaridade entre elas. Para isso, foi utilizada a métrica da Distância de Levenshtein (DL).

Segundo Spinassé e Salgado (2019) a DL é uma forma de mensurar a semelhança entre duas palavras, sendo calculada de acordo com o número de caracteres que necessitam ser alterados para uma palavra chegar à outra. Por exemplo, a DL entre “casa” e “casa” é zero (0), pois nenhum caractere é alterado entre elas. Entre as palavras “casa” e “caza”, a DL é um (1), pois a única alteração foi de “s” para “z”. Já entre as palavras “casa” e “casual” a DL é três (3), porque um caractere foi modificado e outros dois foram adicionados. Sendo assim, quanto menor for a distância de Levenshtein, mais semelhantes são consideradas as palavras, e as variações em questão. Essa etapa corresponde ao passo (4) Figura 1.

Após a execução da etapa 4, o algoritmo realiza o agrupamento para os produtos classificados como semelhantes e realiza a contagem dos produtos mais semelhantes e dos termos mais frequentes. Essa etapa corresponde ao passo (5) Figura 1.

Quanto ao resultado após o processamento realizado foram geradas 2 nuvens de palavras: em uma nuvem são apresentados os nomes de produtos que o cliente e os seus concorrentes oferecem. Enquanto na outra, são apresentados os nomes de produtos que apenas os concorrentes oferecem. Essa etapa corresponde ao passo (6) Figura 1.

Para uma melhorar a visualização do resultado nas nuvens de palavras foi definido um limitador de 100 itens por nuvem.

2.3 Resultados

Após a pesquisa realizada na internet, consultando diversos sites, foram identificadas 304 empresas concorrentes e 7.793 produtos. A Tabela 1 apresenta a quantidade de produtos por UF do corpus textual que foi criado.

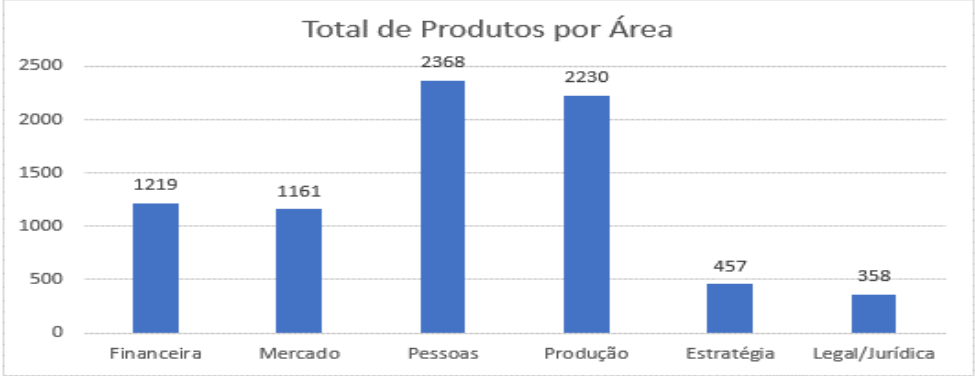
Tabela 1. Total de produtos por UF do corpus textual

Estado	Total de produtos	Estado	Total de produtos
AC	65	PA	161
AL	175	PB	34
AM	64	PE	84
AP	122	PI	177
BA	314	PR	516
CE	230	RJ	321
DF	87	RN	145
ES	99	RO	165
GO	190	RR	113
MA	251	RS	514
MG	413	SC	826
MS	51	SE	147
MT	62	SP	1327
TO	137	Online	1003

Fonte: Dados da pesquisa

A Figura 2 apresenta a quantidade de produtos distribuídos por áreas. Pode-se observar que a maior oferta de produtos ocorre nas áreas de Pessoas e Produção.

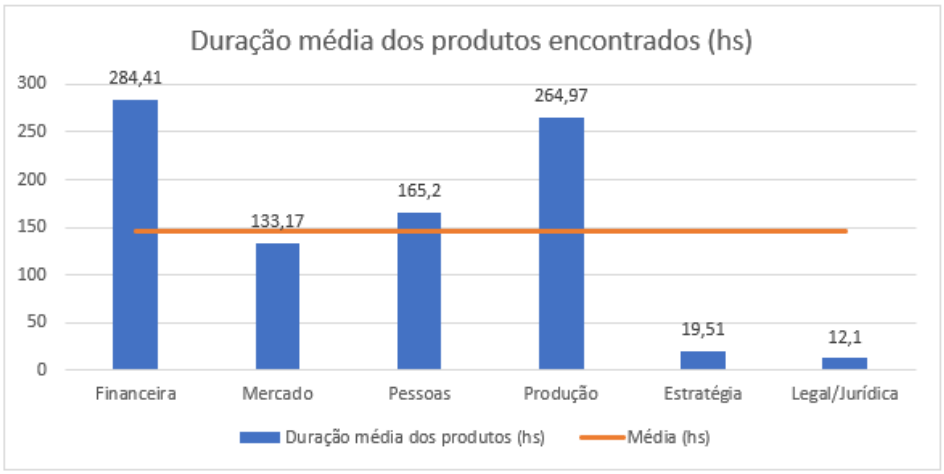
Figura 2. Total de produtos por área



Fonte: Dados da pesquisa

A Figura 3 apresenta a duração média dos produtos encontrados. Os cursos das áreas Financeira e Produção são os mais extensos, com 284,41 e 264,97 horas respectivamente. Enquanto os produtos ofertados na área Legal/Jurídica apresentam a menor duração com 12,1 horas.

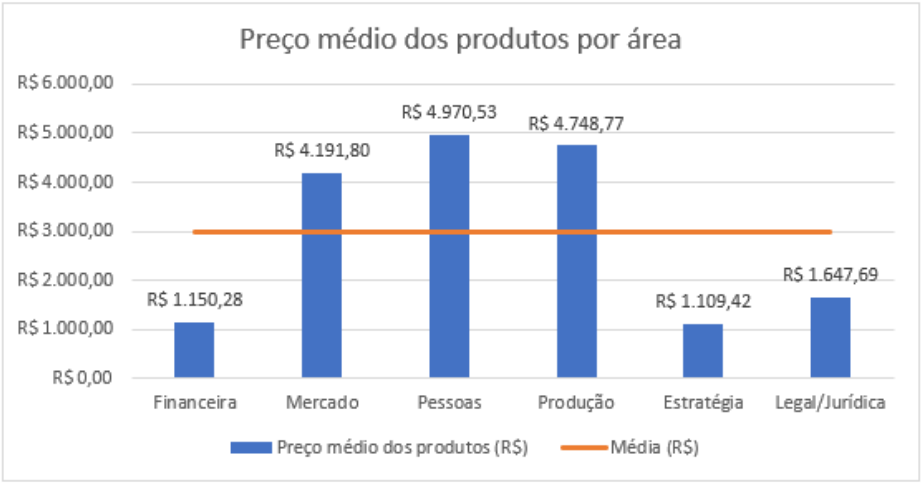
Figura 3. Duração média dos produtos encontrados (hs)



Fonte: Dados da pesquisa

Com relação ao preço dos produtos, a Figura 4 apresenta a média de preços para cada área. Os produtos com valores mais altos são oferecidos nas áreas de Pessoas, Produção e Mercado, com os valores médios de R$ 4.970,53, R$ 4.748,77 e R$ 4.191,80, respectivamente.

Figura 4. Preço médio dos produtos por área

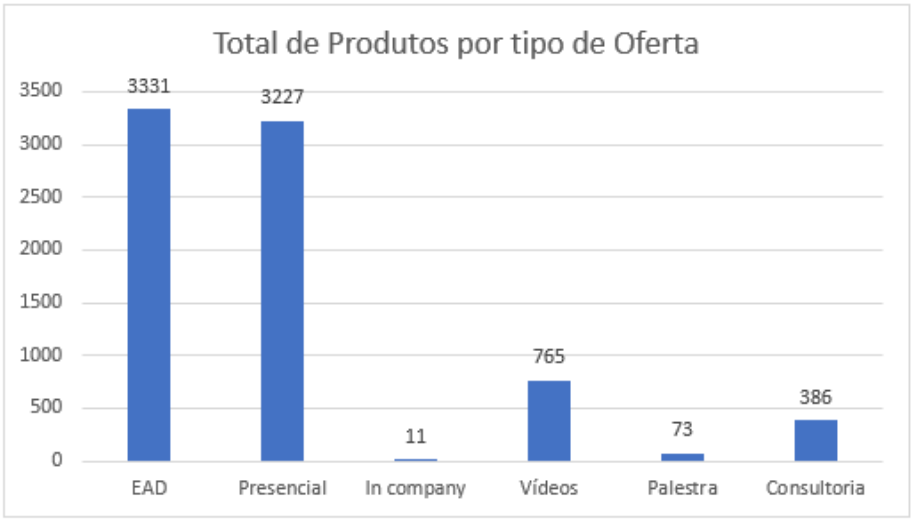


Fonte: Dados da pesquisa

Cabe salientar que essa pesquisa focou nos produtos que foram oferecidos nos últimos dois anos para que as informações de ementa e preço estivessem atualizadas. No entanto, esse período coincide com o atual cenário pandêmico da COVID-19 e era esperado que os cursos ofertados na modalidade EAD ficassem em primeiro lugar. A justificativa para que o segundo lugar ocupado pelos cursos presenciais esteja tão próximo dos cursos EAD é que há cursos que são mistos que foram classificados como presenciais. Eles têm a possibilidade de

estudar a distância, porém tem a necessidade de alguns encontros presenciais. A Figura 5 apresenta a distribuição dos cursos de acordo com o tipo de oferta.

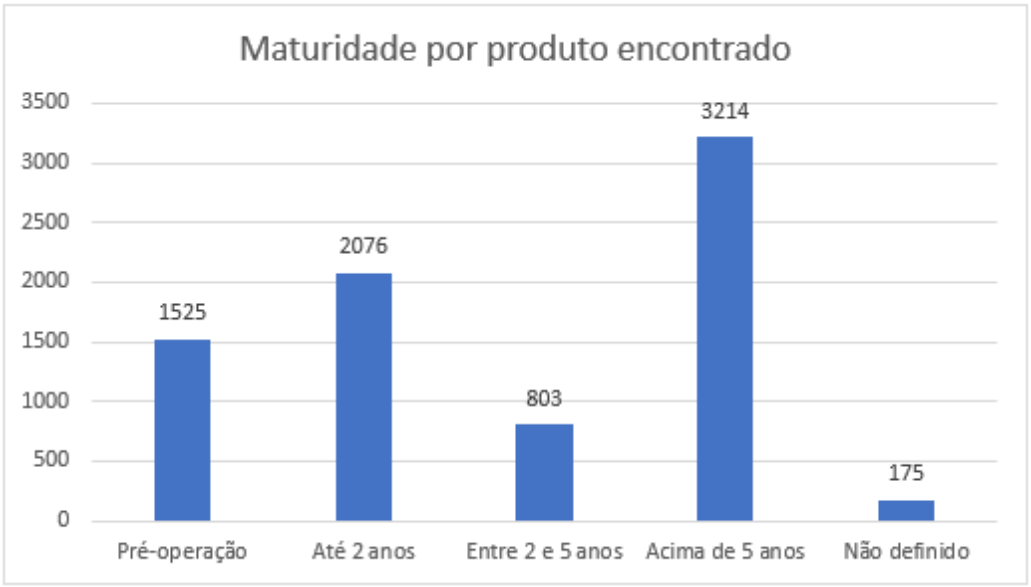
Figura 5. Total de Produtos por tipo de oferta



Fonte: Dados da pesquisa

Com relação a maturidade por produto encontrado apresentada na Figura 6 verifica-se que a maior oferta é de cursos para empresas já consolidadas no mercado, isto é, com mais de 5 anos e fora do período de maior mortalidade das empresas (entre 3 e 5 anos).

Figura 6. Maturidade por produto encontrado



Fonte: Dados da pequisa

As Figuras de 2 a 6 apresentam uma visão geral da base criada como corpus textual desta pesquisa.

A partir da base construída por meio da pesquisa sobre os concorrentes realizada na internet os dados foram normalizados. Logo após, foi criada uma nuvem de palavras que apresenta as palavras encontradas com maior frequência na base. Salientando que as palavras que possuem maior ocorrência são representadas com maior destaque na nuvem. Essas palavras são baseadas no título e descrição dos produtos. Ao analisar os termos mais utilizados pelos concorrentes é possível observar que na oferta atual há um destaque para os conteúdos focados em administração e contabilidade, com destaque na gestão de pessoas, comunicação, projetos, CRM, segmento de alimentos e varejo, e a logística, conforme apresentado na Figura 07.

Figura 7. Nuvem de palavras considerando todos os produtos encontrados na pesquisa.

Fonte: Dados da pesquisa

Ao analisar a área Financeira, assim como nas demais áreas, foi realizado o processo de mineração na descrição e título dos produtos encontrados. Para essa área termos como contabilidade, administração e captação de recursos ganham destaque. Para analisar os dados foram criadas 2 formas: a primeira envolve os produtos que a empresa de capacitação desta pesquisa e suas concorrentes têm em comum, e na segunda, são apresentados os itens que as concorrentes atuam e a empresa de capacitação não é focada.

A Figura 8 apresenta a nuvem de produtos oferecidos pela empresa de capacitação analisada nesta pesquisa com os produtos semelhantes oferecidos pela concorrência.

Figura 8. Nuvem de produtos oferecidos pela empresa de capacitação desta pesquisa com produtos semelhantes aos oferecidos pela concorrência.

Fonte: Dados da Pesquisa

A Figura 09 apresenta a nuvem de produtos que são oferecidos somente pelas empresas concorrentes e não semelhantes aos cursos ofertados pela empresa.

Figura 9. Cursos ofertados somente pela concorrência.

Fonte: Dados da pesquisa

A primeira análise, apresentada nas Figuras 08 e 09, considerou as palavras no contexto de sentença, por exemplo: “Contabilidade com ênfase em Tributos”, e a segunda análise considera as palavras mais utilizadas. A seguir, serão apresentadas as considerações finais da pesquisa realizada.

3 CONSIDERAÇÕES FINAIS

Para viabilizar o desenvolvimento deste trabalho foi necessário criar uma base de dados, coletando as informações definidas para o contexto do problema, em diversos sites na internet. Os dados coletados não estavam estruturados e necessitavam de uma análise a partir do título de cada produto, sendo que, um produto pode ser algum curso, treinamento ou consultoria ofertados na área de capacitação empresarial.

O objetivo principal era identificar quais são os produtos mais ofertados no setor de capacitação empresarial, bem como, mapear quais produtos têm maior peso no mercado e não são oferecidos em uma determinada empresa de capacitação com relação a seus componentes. E partir da extração destas informações foi possível identificar de forma mais clara quais produtos tem bastante ocorrência no mercado (a empresa de capacitação e as concorrentes oferecem o mesmo produto e tem alta incidência), e, identificar aqueles

produtos que estão em alta na concorrência e a empresa de capacitação em questão ainda não oferece.

As técnicas implementadas para minerar o texto e calcular a similaridade entre os termos mostrou-se adequada para nosso objetivo, pois ela permitiu identificar de forma clara quais são os produtos de alta concorrência que são oferecidos tanto pela empresa de capacitação quanto pelos concorrentes, bem como, na identificação de oportunidades de mercado dentro de produtos com grande oferta pelos concorrentes que ainda não são ofertados pela empresa de capacitação.

Sugere-se que a metodologia utilizada seja replicada com uma frequência anual ou semestral, para atualizar a base de dados e realizar novas análises de evolução de tendências de mercado, ou das ofertas de mercado que deixaram de ser relevantes para o público-alvo.

REFERÊNCIAS

ARANHA, Christian; PASSOS, Emmanuel. A Tecnologia de Mineração de Textos. **Revista Eletrônica de Sistemas de Informação**, Curitiba, v. 5, n. 2, p. 1-8, 31 ago. 2006. IBEPES (Instituto Brasileiro de Estudos e Pesquisas Sociais). http://dx.doi.org/10.21529/resi.2006.0502001. Disponível em: http://www.periodicosibepes.org.br/index.php/reinfo/article/view/171. Acesso em: 25 mai. 2022.

BARBOZA, Fabrício Felipe Meleto; FREITAS, Pedro Henrique Chagas. **Modelagem e desenvolvimento de banco de dados**. Porto Alegre: Sagah, 2018.

BARION, Eliana Cristina Nogueira; LAGO, Decio. Mineração de Textos. **Revista de Ciências Exatas e Tecnologia**, Valinhos, v. III , n. 3, p. 123-140, 08 dez. 2008. Disponível em: https://exatastecnologias.pgsskroton.com.br/article/view/2372. Acesso em: 24 mai. 2022.

CHIARELLO, Filippo; FANTONI, Gualtiero; HOGARTH, Terence; GIORDANO, Vito; BALTINA, Liga; SPADA, Irene. Towards ESCO 4.0 – Is the European classification of skills in line with Industry 4.0? A text mining approach. **Technological Forecasting And Social Change**, Reino Unido, v. 173, p. 1-18, 09 set. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.techfore.2021.121177. Disponível em: https://www.sciencedirect.com/science/article/pii/S0040162521006107. Acesso em: 25 mai. 2022

CHUNG, Chih-Hung; CHEN, Lu-Jia. Text mining for human resources competencies: Taiwan example. **European Journal Of Training And Development**, Washington, v. 45, n. 6/7, p. 588-602, 16 set. 2021. Emerald. http://dx.doi.org/10.1108/ejtd-07-2018-0060. Disponível em: https://eric.ed.gov/?id=EJ1308738. Acesso em: 25 mai. 2022.

CORDELLA, Barbara; GRECO, Francesca; MEOLI, Paolo; PALERMO, Vittorio; GRASSO, Massimo. Educational Culture and Job Market: a text mining approach. **Studies In Classification, Data Analysis, And Knowledge Organization**, [s. l], p. 287-297, 25 nov. 2020. Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-52680-1_23. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-52680-1_23. Acesso em: 25 mai. 2022

GALATI, Francesco; BIGLIARDI, Barbara. Industry 4.0: emerging themes and future research avenues using a text mining approach. **Computers In Industry**, Tarbes, v. 109, p. 100-113, 06 maio 2019. Elsevier BV. http://dx.doi.org/10.1016/j.compind.2019.04.018. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0166361518307772. Acesso em: 20 mai. 2022

GONÇALVES, Eduardo Corrêa. **Mineração de texto - Conceitos e aplicações práticas**. 2012. Disponível em: https://www.devmedia.com.br/mineracao-de-texto-conceitos-e-aplicacoes-praticas-revista-sql-magazine-105/26328. Acesso em: 20 maio 2022.

IGNOATTO, Maicon Luiz; WEBBER, Carine Geltrudes. Inteligência Competitiva nas Mídias Sociais: um estudo de caso na moda. **Scientia Cum Industria**, Caxias do Sul, v. 7, n. 2, p. 156-164, 26 nov. 2019. Universidade Caxias do Sul. http://dx.doi.org/10.18226/23185279.v7iss2p156. Disponível em: http://www.ucs.br/etc/revistas/index.php/scientiacumindustria/article/view/7749/3980. Acesso em: 20 mai. 2022.

KUMAR, Sunil; KAR, Arpan Kumar; ILAVARASAN, P. Vigneswara. Applications of text mining in services management: a systematic literature review. **International Journal Of Information Management Data Insights**, Delhi, v. 1, n. 1, p. 1-14, 26 fev. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.jjimei.2021.100008. Disponível em: https://www.sciencedirect.com/science/article/pii/S266709682100001X. Acesso em: 20 mai. 2022.

SPINASSÉ, Karen Pupp; SALGADO, Bruna Miskinis. Pesquisando a inteligibilidade entre o Hunsrückisch e o alemão standard. **Contingentia**, Porto Alegre, v. 7, n. 1, p. 9-28, jun. 2019. Disponível em: https://lume.ufrgs.br/handle/10183/200788. Acesso em: 26 mai. 2022.

LEON, Linda A.; SEAL, Kala Chand; PRZASNYSKI, Zbigniew H.; WIEDENMAN, Ian. Skills and competencies required for jobs in business analytics: A content analysis of job advertisements using text mining. **Operations And Service Management**, [S.L.], p. 880-904, 2018. IGI Global. http://dx.doi.org/10.4018/978-1-5225-3909-4.ch041.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L.. **Mineração de Textos**. Goiás: Instituto de Informática, 2007. 30 p. INF_005/07. Disponível em: http://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf. Acesso em: 20 mai. 2022.

NOTA, Giancarlo; POSTIGLIONE, Alberto; CARVELLO, Rosario. Text mining techniques for the management of predictive maintenance. **Procedia Computer Science**, Salerno, v. 200, p.

778-792, 2022. Elsevier BV. http://dx.doi.org/10.1016/j.procs.2022.01.276. Disponível em: https://www.sciencedirect.com/science/article/pii/S187705092200285X. Acesso em: 20 ago. 2022