

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA QUÍMICA E ENGENHARIA DE ALIMENTOS

Alice Beal Kerber

**TENDÊNCIAS NAS PUBLICAÇÕES RELACIONADAS COM A INDÚSTRIA DE
ALIMENTOS**

Florianópolis

2022

Alice Beal Kerber

**TENDÊNCIAS NAS PUBLICAÇÕES RELACIONADAS COM A INDÚSTRIA DE
ALIMENTOS**

Trabalho Conclusão do Curso de Graduação em Engenharia de Alimentos do Departamento de Engenharia Química e de Alimentos do Centro Tecnológico da Universidade Federal de Santa Catarina, como requisito para a obtenção do título de Bacharel em Engenharia de Alimentos.
Orientador: Prof. Dr. Natan Padoin

Florianópolis

2022

AGRADECIMENTOS

Aos professores, servidores e colegas do departamento de engenharia química e de alimentos da UFSC. Ao Prof. Dr. Natan Padoin pelo apoio, confiança e pela liberdade em conduzir este trabalho. Aos membros da banca, Prof.^a Dr.^a Cíntia Soares e Prof. Dr. Sergio Yesid Gomes Gonzales, por terem aceitado o convite.

Agradeço a todas as pessoas que deram suporte durante este trabalho e durante toda a graduação, incluindo minha família e amigos. Dessas pessoas queridas, um agradecimento especial para Jessica Daiane Domingos pelo apoio incondicional.

Agradeço a GeekHunter e todas as pessoas que trabalhei em conjunto, pelos aprendizados e por terem aberto a porta para o universo de dados.

RESUMO

Escolher qual será a próxima pesquisa que agregará valor à indústria de alimentos nem sempre é fácil, mas a identificação de tendências pode direcionar essa escolha. Este trabalho se propõe a, através da identificação dos tópicos mais frequentes dos documentos da base Scopus, apontar tendências e oportunidades nos conteúdos relacionados com a indústria de alimentos. A metodologia CRISP-DM foi utilizada neste trabalho, e o algoritmo Latent Dirichlet Allocation foi aplicado para a identificação dos tópicos. Observou-se que os últimos 21 anos concentram a maior parte dos documentos publicados, o que corresponde a cerca de 88% de todos os documentos utilizados neste trabalho. Com a extração de tópicos, foi possível perceber como as temáticas de secagem, segurança alimentar, atividade antioxidante e crescimento bacteriano são frequentes e consequentemente relevantes para a indústria de alimentos.

Palavras-chave: Tendências. Engenharia de alimentos. Indústria de alimentos. Extração de tópicos.

ABSTRACT

Choosing the next research that will add value to the food industry is not always easy, but the identification of trends can guide this choice. This work proposes, through the identification of the most frequent topics in Scopus documents, to point out trends and opportunities in the contents related to the food industry. The CRISP-DM methodology was used in this work. Moreover, the Latent Dirichlet Allocation algorithm was applied to identify the topics. It was noticed that the last 21 years concentrate most of the published documents, which is about 88% of all documents used in this work. With the extraction of topics, it was possible to notice how the themes of drying, food safety, antioxidant activity and bacterial growth are frequent and consequently relevant to the food industry.

Keywords: Trends. Food Engineering. Food industry. Topic extraction.

LISTA DE FIGURAS

Figura 1 - Opções de exportação.....	13
Figura 2 - Exemplo de workflow utilizado neste trabalho.....	17
Figura 3 - Modelo gráfico para LDA.....	25
Figura 4 - Configurações padrão do nó <i>Topic Extractor (Parallel LDA)</i>	27
Figura 5 - 20 áreas do conhecimento com mais documentos.....	44
Figura 6 - Resultados por ano de publicação.....	45
Figura 7 - 20 áreas do locais com mais documentos.....	47
Figura 8 - Proporção dos documentos por tipo.....	48
Figura 9 - 20 instituições com mais documentos.....	48
Figura 10 - 20 fontes com mais documentos.....	49
Figura 11 - 20 financiadores com mais documentos.....	50
Figura 12 - 20 pessoas autoras com mais documentos.....	50

LISTA DE QUADROS

Quadro 1 - Descrição das variáveis normalmente utilizadas	23
Quadro 2 - Teste de extração de tópicos dos títulos.....	24
Quadro 3 - Comparação tópicos para variações dos parâmetros.....	27
Quadro 4 - Teste quantidades de palavras por tópico.....	29
Quadro 5 - Teste de quantidade de tópicos - Período inteiro.....	32
Quadro 6 -Teste de quantidade de tópicos - 1905 - 1999.....	35
Quadro 7 - Teste de quantidade de tópicos - 2000 - 2010.....	37
Quadro 8 - Teste de quantidade de tópicos - 2011 - 2021.....	39
Quadro 9 - Teste de quantidade de tópicos de afiliações brasileiras.....	41
Quadro 10 - Resultados da busca intermediária por área do conhecimento....	42
Quadro 11 - Métricas descritivas dos 4 recortes de tempo.....	45
Quadro 12 - Resultados da pesquisa utilizada na introdução.....	58

LISTA DE TABELAS

Tabela 1 - Resultados da busca intermediária por área do conhecimento	43
Tabela 2 - Métricas descritivas dos 4 recortes de tempo.....	46
Tabela 3 - Quantidade de documentos atribuídos aos tópicos selecionados...	53
Tabela 4: Quantidade de documentos atribuídos aos tópicos selecionados de afiliações brasileiras.....	55

LISTA DE ABREVIATURAS E SIGLAS

CRISP-DM: *Cross-Industry Standard Process for Data Mining*

LDA: *Latent Dirichlet Allocation*

SUMÁRIO

INTRODUÇÃO	11
OBJETIVOS	11
Objetivo Geral	11
Objetivos Específicos	11
MATERIAIS E MÉTODOS	13
DADOS	13
Pesquisa	14
KNIME Analytics Platform	17
CRISP-DM	17
Compreensão do negócio	18
Compreensão dos dados	21
Preparação dos dados	21
Modelagem	22
Avaliação	45
Implementação	45
RESULTADOS E DISCUSSÃO	46
ANÁLISE MACRO	46
Áreas do conhecimento	46
Evolução da quantidade nos anos	48
Documentos por local	50
Documentos por tipo	50
Documentos por afiliação	51
Documentos por fonte	52
Documentos por financiamento	52
Documentos por autoria	53
TÓPICOS	53
Tópicos de afiliações brasileiras	57
CONCLUSÃO	59
REFERÊNCIAS	60
APÊNDICE	62

1. INTRODUÇÃO

Escolher qual será a próxima pesquisa nem sempre é fácil, mas a identificação de tendências pode direcionar essa escolha. Ao pesquisar na base Scopus pelas palavras *food engineering trends* e *food industry trends*, os primeiros resultados (ordenando por publicações mais recentes) não retornam um estudo abrangente na avaliação de tendências. A publicação abrangente mais recente com resumo e texto disponíveis é de 2011: *The Beginning, Current, and Future of Food Engineering: A Perspective* (HELDMAN; LUND, 2011). Outros resultados encontrados costumam abordar uma temática específica (*nicho*) dentro da engenharia de alimentos ou indústria de alimentos. A ferramenta *Google Scholar* também foi utilizada e apenas um resultado foi considerado abrangente e recente: *Challenges, trends and opportunities in food processing* (AUGUSTO, 2020). Porém, utiliza uma abordagem diferente da proposta neste trabalho. Os resultados dessas buscas podem ser conferidos no apêndice. O desejo de uma avaliação global sobre as tendências na Engenharia de Alimentos impulsiona este trabalho.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Através da identificação dos tópicos mais frequentes e a distribuição dos documentos, apontar tendências e oportunidades nos conteúdos relacionados com a indústria de alimentos.

1.1.2 Objetivos Específicos

- Realizar uma análise macro dos documentos disponíveis na base Scopus.
- Identificar os tópicos mais frequentes em todo o período disponível.
- Observar a evolução dos tópicos nos períodos escolhidos.
- Identificar países e instituições de ensino que produzem mais documentos.

- Comparar os tópicos mundiais com os tópicos de publicações brasileiras.

2. MATERIAIS E MÉTODOS

2.1 DADOS

Para a realização das análises, dois critérios foram considerados no momento de escolha da base de dados: qualidade e quantidade. Sem dados com qualidade, não é possível realizar uma boa análise e chegar em um resultado coerente. Sem um volume mínimo, os resultados não serão representativos, podendo ocorrer um *overfit* do modelo criado.

Scopus é o maior banco de dados de resumos e citações da literatura com revisão por pares: revistas científicas, livros, produção oriunda de congressos, entre outras. Os documentos e suas informações complementares são padronizadas, o que facilita a análise. Scopus possui uma curadoria que assegura a qualidade de sua base. No repositório de conteúdo, são 3,7 TB de dados armazenados, 1,4 bilhão de referências, 70.000 perfis institucionais e 16 milhões de perfis de autores (ELSEVIER, [s.d.]).

Além disso, a base de dados permite a exportação das palavras chaves e resumos, o que possibilita uma comparação entre eles.

Figura 1 - Opções de exportação.

Export document settings ⓘ

You have chosen to export 29239 documents

Select your method of export

MENDLEY ExLibris RefWorks RIS Format EndNote, Reference Manager CSV Excel BibTeX Plain Text ASCII in HTML

What information do you want to export?

<input type="checkbox"/> Citation information	<input type="checkbox"/> Bibliographical information	<input type="checkbox"/> Abstract & keywords	<input type="checkbox"/> Funding details	<input type="checkbox"/> Other information
<input type="checkbox"/> Author(s)	<input type="checkbox"/> Affiliations	<input type="checkbox"/> Abstract	<input type="checkbox"/> Number	<input type="checkbox"/> Tradenames & manufacturers
<input type="checkbox"/> Author(s) ID	<input type="checkbox"/> Serial identifiers (e.g. ISSN)	<input type="checkbox"/> Author keywords	<input type="checkbox"/> Acronym	<input type="checkbox"/> Accession numbers & chemicals
<input type="checkbox"/> Document title	<input type="checkbox"/> PubMed ID	<input type="checkbox"/> Index keywords	<input type="checkbox"/> Sponsor	<input type="checkbox"/> Conference information
<input type="checkbox"/> Year	<input type="checkbox"/> Publisher		<input type="checkbox"/> Funding text	<input type="checkbox"/> Include references
<input type="checkbox"/> EID	<input type="checkbox"/> Editor(s)			
<input type="checkbox"/> Source title	<input type="checkbox"/> Language of original document			
<input type="checkbox"/> volume, issue, pages	<input type="checkbox"/> Correspondence address			
<input type="checkbox"/> Citation count	<input type="checkbox"/> Abbreviated source title			
<input type="checkbox"/> Source & document type				
<input type="checkbox"/> Publication Stage				
<input type="checkbox"/> DOI				
<input type="checkbox"/> Open Access				

Fonte: (Scopus, 2022)

2.1.1 Pesquisa

Com a finalidade de evitar comparar anos concluídos com um ano em andamento, optou-se por extrair publicações posteriores a 2021. Foram testadas diversas formas de busca: com palavras-chave diferentes, com a restrição de estarem lado a lado, sem essa restrição, com certa distância entre as palavras, apenas o radical de tal palavra, excluindo certas áreas do conhecimento, etc.

Após os testes da forma de busca, foram identificados os operadores e códigos mais adequados para os objetivos do trabalho, sendo esses:

- TITLE-ABS-KEY: um campo combinado que pesquisa resumos, palavras-chaves e títulos de documentos. Ao colocar *food*, por exemplo, os documentos retornados terão a palavra *food* nos seus resumos, títulos de artigos ou campos de palavras-chaves.
- * : a palavra pode ter mais letras depois do * e continuará sendo um resultado válido. Exemplo: *engineer** retorna *engineering* também.
- w/10: distância entre as palavras de até 10 palavras, independente da ordem. Scopus recomenda o número 15 como um equivalente de buscar por palavras na mesma frase, mas optou-se por uma distância mais conservadora.
- PUBYEAR: ano de publicação.
- OR: operador “ou”, o resultado precisa ter tal condição ou a outra. Exemplo: *food engineering OR food industry* o resultado vai ter as palavras *food engineering* ou *food industry*.
- AND: operador “e”, o resultado precisa ter tal condição e a outra. Exemplo: *food AND engineering* o resultado vai ter ambas as palavras.
- LIMIT-TO: refinar a pesquisa em tal critério.
- SUBJAREA: *subject area*, área do conhecimento. O mesmo documento pode ser categorizado em mais de uma área.
- "AGRI": *Agricultural and Biological Sciences*
- "ENGI": *Engineering*
- "BIOC": *Biochemistry, Genetics and Molecular Biology*
- "CHEM": *Chemistry*
- "CENG": *Chemical Engineering*

- "IMMU": *Immunology and Microbiology*
- "PHAR": *Pharmacology, Toxicology and Pharmaceutics*
- "MATE": *Materials Science*
- "COMP": *Computer Science*
- "PHYS": *Physics and Astronomy*
- "MATH": *Mathematics*
- AFFILCOUNTRY: A parte do país de um endereço de afiliação da pessoa autora. Foi utilizado para filtrar os documentos de brasileiros.

Realizou-se uma busca intermediária, antes de serem escolhidas as áreas do conhecimento que mais se enquadram na temática do trabalho:

TITLE-ABS-KEY

```
(
(food w/10 engineer*)
OR (food w/10 industry)
OR (food w/10 process*)
OR (food w/10 tech*)
OR (food w/10 science)
OR (food w/10 manufact*)
OR (food w/10 product*)
OR (food w/10 develop*)
    OR (food w/10 innovat*)
)
```

AND PUBYEAR < 2022

A partir do resultado dessa busca, foram avaliadas cada área do conhecimento e seus resultados. Para as amostras, foi escolhido um grau de confiança de 80% e margem de erro de 10%, apesar da prática mais comum ser 95% e 5%, respectivamente, pois seria inviável avaliar a quantidade de documentos necessária para uma assertividade maior. Para cada área, foram analisados 41 documentos coletados de forma aleatória. Apenas documentos com *abstract* foram utilizados. Para identificar se tal documento é relacionado com a temática, foram utilizados os campos em ordem de relevância: *Title, Abstract, Author Keywords, Index*

Keywords, Affiliations, Source title, Publisher. Para simplificar a avaliação, o principal fator avaliado foi se tal documento tem relação com produção de alimentos em escala industrial. Se 70% dos documentos avaliados de determinada área foram considerados como relacionados, essa área foi utilizada na busca final.

A consulta final escolhida nesta pesquisa foi, então:

TITLE-ABS-KEY

```
(  
  (food w/10 engineer*)  
  OR (food w/10 industry)  
  OR (food w/10 process*)  
  OR (food w/10 tech*)  
  OR (food w/10 science)  
  OR (food w/10 manufact*)  
  OR (food w/10 product*)  
  OR (food w/10 develop*)  
    OR (food w/10 innovat*)  
)  
AND PUBYEAR < 2022
```

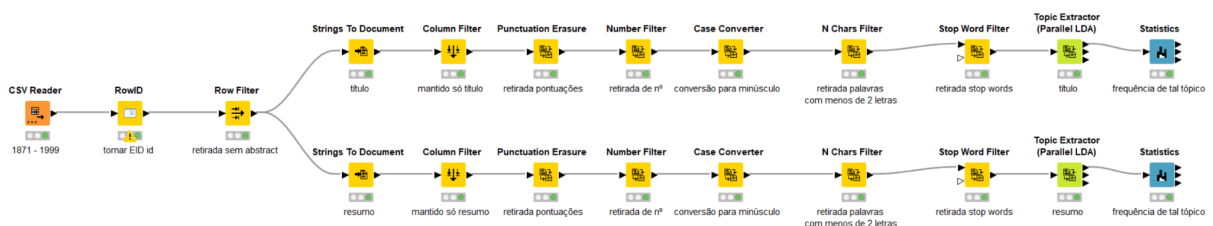
AND

```
(  
  LIMIT-TO ( SUBJAREA,"AGRI" )  
  OR LIMIT-TO ( SUBJAREA,"ENGI" )  
  OR LIMIT-TO ( SUBJAREA,"BIOC" )  
  OR LIMIT-TO ( SUBJAREA,"CHEM" )  
  OR LIMIT-TO ( SUBJAREA,"CENG" )  
  OR LIMIT-TO ( SUBJAREA,"IMMU" )  
  OR LIMIT-TO ( SUBJAREA,"PHAR" )  
  OR LIMIT-TO ( SUBJAREA,"MATE" )  
  OR LIMIT-TO ( SUBJAREA,"COMP" )  
  OR LIMIT-TO ( SUBJAREA,"PHYS" )  
  OR LIMIT-TO ( SUBJAREA,"MATH" )  
)
```


2.2 KNIME Analytics Platform

A leitura dos arquivos, o tratamento dos dados e a extração de tópicos foram realizados na plataforma analítica KNIME, um software de código aberto em ciência de dados. Nesta ferramenta foram construídos os fluxos de dados. A figura a seguir ilustra um desses fluxos, desde a leitura dos arquivos até a extração de tópicos.

Figura 2: Exemplo de workflow utilizado neste trabalho.



Fonte: a autora (2022).

2.3 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) é conhecido como uma metodologia ou processo para projetos de *Data Science*, sendo o mais conhecido para tal propósito. Foi criado por DaimlerChrysler, SPSS e NCR. Ele consiste em seis etapas:

- *Business understanding* (Compreensão do negócio) - esta fase inicial foca em entender os objetivos do projeto e requisitos de uma perspectiva de negócios, então convertendo este conhecimento na definição de um problema de mineração de dados e um plano preliminar desenhado para alcançar os objetivos.
- *Data understanding* (Compreensão dos dados) - esta fase inicia com uma coleta inicial de dados e prossegue com atividades com o objetivo de se familiarizar com os dados, para identificar problemas de qualidade dos dados, para descobrir os primeiros *insights* dos dados ou para detectar subconjuntos interessantes para formular hipóteses para informações ocultas.

- *Data preparation* (Preparação dos dados) - esta fase cobre todas as atividades para construir o conjunto de dados final a partir dos dados brutos.
- *Modeling* (Modelagem) - nesta fase, várias técnicas de modelagem são selecionadas e aplicadas e os parâmetros são calibrados para valores ideais.
- *Evaluation* (Avaliação) - neste estágio o modelo (ou modelos) obtidos são avaliados de forma mais aprofundada e as etapas executadas para construir o modelo são revisadas para garantir que atinja os objetivos de negócio de forma adequada.
- *Deployment* (Implementação) - a criação do modelo normalmente não é o fim do projeto. Mesmo quando o propósito do modelo é aumentar o conhecimento dos dados, o conhecimento adquirido precisará ser organizado e apresentado de forma que o cliente possa usá-lo (AZEVEDO; SANTOS, [s.d.]).

2.3.1 Compreensão do negócio

A introdução e os objetivos no início deste trabalho representam parte da etapa *Business understanding* do processo CRISP-DM. Portanto, não serão repetidos aqui. Além do contexto inicial e objetivos, nessa etapa foram avaliados riscos, restrições e dependências.

Um risco neste trabalho são os falsos positivos, quando documentos que não são relacionados com a indústria de alimentos são retornados nas buscas. Para minimizar esse risco, foram realizados testes com buscas diferentes para identificar qual configuração de busca trazia a maior quantidade de documentos não relacionados. Foi escolhido ordenar por relevância por se tratar justamente dos resultados que deveriam ser mais relacionados com as palavras utilizadas na busca.

O mecanismo de pesquisa da Scopus usa um modelo de relevância baseado em conceitos da ciência de recuperação de informações, longa experiência com pesquisa na web, dados e empresas. Cada documento retornado como resultado de uma busca recebe uma pontuação composta por diversos fatores. O resultado do cálculo depende da própria consulta, dos documentos pesquisados e da maneira que esses dois fatores são equilibrados entre si. São considerados os seguintes fatores: frequência dos termos no documento; quão significativa é a palavra (*document frequency* TF/IDF conceito introduzido por Karen Spärck Jones); onde a palavra foi encontrada (título, resumo, palavras chaves); posição no documento (quanto mais

cedo, provavelmente é mais importante); proximidade entre as palavras da busca; completude (se o documento tiver todas as palavras da busca no mesmo campo recebe uma nota mais alta) (“What does ‘Relevance’ mean in Scopus? - Scopus: Access and use Support Center”, [s.d.]). Foram avaliados de forma manual os 20 resultados com maior relevância para cada busca, observando título, resumo e palavras-chave. Durante as avaliações, percebeu-se que nem sempre é fácil traçar a linha do quão relacionado um artigo pode ser com a temática de indústria de alimentos. Só houve uma exceção que levou segundos para se perceber que não era relacionado: um estudo realizado com estudantes de engenharia de alimentos relacionado com o nível de inglês dessas pessoas. Os demais documentos avaliados de fato apresentavam algum nível de relação com a temática, ou então com fundamentos da área.

A partir desse cenário, foram testadas duas formas para diminuir os falsos positivos: limitando a certas áreas relacionadas com a temática ou excluindo certas áreas sem relação. Durante a realização dos testes para reduzir os falsos positivos, percebeu-se que o mesmo documento pode ser categorizado em mais de uma área do conhecimento, o que significa que um documento que faz parte de uma área que intuitivamente não tem relação com a indústria de alimentos também pode estar em outra categoria relacionada. Por exemplo, um documento pode estar categorizado em *Arts and Humanities* e em *Chemical Engineering*. Optando por excluir a área *Arts and Humanities*, tal documento seria retirado dos resultados, o que geraria um declínio no volume da amostra utilizada para este trabalho. Para evitar desconsiderar documentos relevantes, foi escolhida a opção “limitar a certas áreas relacionadas”, ao invés de excluir tais áreas não relacionadas. A lista final de áreas foi detalhada anteriormente na seção “Pesquisa” deste trabalho.

Outra possibilidade considerada para reduzir a ocorrência de falsos positivos foi excluir documentos com determinadas palavras-chave. Embora seja uma abordagem bastante direta, considerou-se o fato de que a escolha das palavras-chave de um artigo depende somente do(s) autor(es), podendo uma seleção de documentos ou artigos não seguir a mesma lógica de atribuição. Nas avaliações manuais, por exemplo, foram encontrados alguns documentos com temática voltada para a área de educação, que não estavam diretamente relacionados com processamento de alimentos. Considerou-se descartar esse tipo de resultado, mas nem todo documento pertencente à área de educação apresentava palavras-chave como "educação". Além

disso, não é tão incomum um documento não apresentar palavras-chave. Da amostra de documentos utilizada na extração de tópicos de todo o período, 26% não possuem *Author Keywords* e 25% não possuem *Index Keywords*.

Percebeu-se que ambas as alternativas avaliadas apresentavam um problema em comum: a possibilidade de descartar documentos relevantes de todas as outras etapas seguintes do trabalho por uma intervenção na forma da pesquisa. Ou seja, ao tentar reduzir o risco de falsos positivos, gerar-se-ia outro risco. Por isso, a estratégia escolhida foi reduzir o risco de falsos positivos com o mínimo de intervenção possível, com uma escolha cuidadosa de palavras. Nesse caso, foi adicionada uma condição para considerar apenas as palavras escolhidas em uma mesma frase. Além disso, ao invés de descartar o documento cuja condição fosse atendida, optou-se por limitar a determinadas áreas. Caso a quantidade de falsos positivos seja grande o suficiente para interferir na extração de tópicos, o resultado deve ser refinado, com a retirada dos tópicos não relacionados com a indústria de alimentos.

Este trabalho depende das informações que foram fornecidas na base Scopus. Sendo assim, se certo campo não foi preenchido, não existe a possibilidade de solicitar o preenchimento ou adição de informações pelos autores. Além da presença ou não de tal dado, a clareza do que está disponível também possui essa dependência. Na hora de escolher quais áreas traziam resultados relacionados com a indústria de alimentos, por exemplo, foram revisados certos resumos e títulos que não eram explicativos o suficiente para avaliar a relação. Outro exemplo é a linguagem do resumo. Apesar do padrão internacional ser a língua inglesa, foram encontrados documentos em outra língua em quantidade o suficiente para interferir nos tópicos extraídos do período de 1905 até 1999.

É válido ressaltar também uma restrição na pesquisa na base Scopus: não é possível restringir a pesquisa para retornar apenas documentos que possuem um determinado campo não vazio, especialmente no caso do *abstract*. Isso gerou um esforço maior no tratamento dos dados, e apenas depois desta etapa foi possível determinar o número total de documentos que poderiam seguir para a extração de tópicos.

2.3.2 Compreensão dos dados

Nesta etapa, realizou-se uma análise macro, cujo resultado será detalhado na parte de resultados e discussão. Percebeu-se que determinados documentos não possuem resumo. Portanto, tais documentos não serão utilizados na extração de tópicos. A pesquisa foi refinada nesta etapa também.

2.3.3 Preparação dos dados

Os resultados da pesquisa foram exportados do banco de dados Scopus no formato .csv. Em seguida, realizou-se a leitura e conferência desses arquivos, uma etapa importante para a qualidade dos dados. Quatro fatores provocam erros na leitura, resultando em linhas que não distribuíram os dados nas colunas da forma correta: *[No abstract available]*, *[No author name available]*, *[No author id available]* e aspas no meio do texto. As palavras que correspondiam a tal campo com um desses erros ficavam espalhadas nas colunas incorretas, o que interferia na distribuição dos campos seguintes.

No caso dos documentos sem *abstract*, o tratamento foi deletar tal linha, pois o resumo é um campo obrigatório para este trabalho. As aspas no meio do texto foram retiradas de forma manual e as palavras foram colocadas na coluna correta. As informações correspondentes a *[No author name available]* e *[No author id available]* foram rearranjadas nas colunas correspondentes. Esses tratamentos e a identificação dos erros foram realizados com o auxílio da ferramenta LibreOffice Calc.

Durante o tratamento de dados, o estágio de pré-processamento geralmente contém tarefas como tokenização, remoção de *stop-words*, conversão para minúsculas e lematização. O estágio de extração de características geralmente utiliza o modelo de espaço vetorial que faz uso da abordagem *bag-of-words*. O estágio de seleção de recursos, na maioria das vezes, emprega os métodos de filtragem, como a frequência do documento (UYSAL; GUNAL, 2014).

No processamento de texto, a tokenização é o procedimento de dividir um texto em palavras, frases ou outras partes significativas, conhecidas como tokens. Em outras palavras, a tokenização é uma forma de segmentação de texto. Normalmente, a segmentação é realizada considerando apenas caracteres alfabéticos ou alfanuméricos que são delimitados por caracteres não alfanuméricos (por exemplo,

pontuações, espaços em branco). Foi utilizado o nó *Strings To Document* para fazer a tokenização com *OpenNLP SimpleTokenizer*.

Alguns tratamentos pequenos foram realizados nos dados: remoção das colunas que não seriam utilizadas na extração de tópicos; retirada dos caracteres de pontuação com o nó *Punctuation Erasure*; exclusão dos termos representando números com o nó *Number Filter*; e remoção de palavras com menos de três caracteres.

Uma etapa relevante foi a retirada de *Stop-words*, que são as palavras comumente encontradas em textos sem dependência de um tópico específico (por exemplo, conjunções, preposições, artigos, etc.). Portanto, as palavras são geralmente consideradas irrelevantes em estudos de classificação de texto e removidas antes da classificação. As palavras irrelevantes são específicas para o idioma que está sendo estudado, assim como no caso de lematização. Foi utilizado o nó *Stop Word Filter* configurado para *English*, pois os títulos e resumos em geral estão em inglês.

Outra etapa de pré-processamento amplamente usada para classificação de texto é a conversão para letras minúsculas. Uma vez que as formas das palavras em maiúsculas ou minúsculas são consideradas sem diferença, todos os caracteres em maiúsculas são geralmente convertidos para suas formas em minúsculas antes da classificação. Existem algoritmos e ferramentas que trabalham melhor com todas as palavras minúsculas, por isso é uma prática comum. Foi utilizado o nó *Case Converter* configurado para *To lower case*.

2.3.4 Modelagem

Na etapa de modelagem, aplicou-se o algoritmo *Latent Dirichlet Allocation* para a identificação dos tópicos. Foram testados valores diferentes dos parâmetros alfa e beta. LDA é um modelo probabilístico generativo para dados discretos, como *bag of words* de documentos de texto. Neste contexto, esse modelo é frequentemente chamado de modelo de tópicos.

Quadro 1 - Descrição das variáveis normalmente utilizadas.

Variável	Descrição
D	Número de documentos na coleção
W	Número de palavras distintas no vocabulário
N	Número total de palavras na coleção
K	Número de tópicos
x_{ij}	i^{th} palavra observada no documento j
z_{ij}	Tópico atribuído x_{ij}
N_{wk}	Contagem de palavras atribuídas pro tópico
N_{kj}	Contagem de tópicos atribuídos em um documento
ϕ_k	Probabilidade de palavra dado tópico k
θ_j	Probabilidade de tópico dado documento j

Fonte: (NEWMAN et al., [s.d.]).

O LDA modela cada um dos documentos D em uma coleção como uma mistura de K tópicos latentes, com cada tópico sendo uma distribuição multinomial sobre um vocabulário de W palavras. Para cada documento j , primeiro extraímos uma porção de mistura θ_j de um Dirichlet com um parâmetro alfa (α). Para a palavra i^{th} no documento, um tópico $z_{ij} = k$ é extraído com a probabilidade de $\theta_{k|j}$. A palavra x_{ij} é então retirada do tópico z_{ij} , com x_{ij} tomando no valor w com probabilidade $\theta_{w|z_{ij}}$. Um Dirichlet anterior com parâmetro beta (β) é colocado nas distribuições *word-topic* ϕ_k . Assim, o processo generativo para LDA é dado por:

$$\theta_j \sim D[\alpha], \quad \phi_k \sim D[\beta], \quad z_{ij} \sim \theta_j, \quad x_{ij} \sim \phi_{z_{ij}} \quad (1)$$

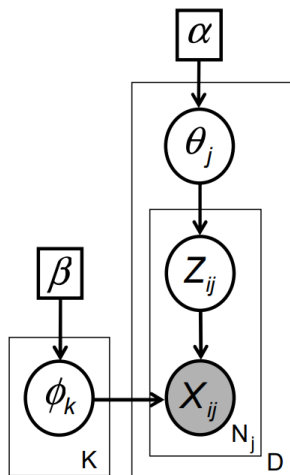
Para evitar desordem, denotamos a amostragem de um Dirichlet $\theta_j \sim D[\alpha]$ como uma abreviação de $[\theta_{1|j}, \dots, \theta_{K|j}] \sim D[\alpha, \dots, \alpha]$, e da mesma forma para ϕ .

Dirichlet anteriores simétricos são utilizados pela simplicidade. A distribuição completa juntando todos os parâmetros e variáveis é:

$$p(x, z, \theta, \phi | \alpha, \beta) = \prod_j \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_k \theta_{kj}^{\alpha} \prod_k \frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_w \phi_{wk}^{N_{wk} + \beta - 1} \quad (2)$$

onde $N_{wkj} = \#\{i: x_{ij} = w, z_{ij} = k\}$, e usamos a convenção que os índices faltantes são resumidos. $N_{kj} = \sum_w N_{wkj}$ e $N_{wk} = \sum_j N_{wkj}$ são as duas matrizes de contagem primárias usadas nos cálculos, representando o número de palavras atribuídas ao tópico k no documento j , e o número de vezes que a palavra w é atribuída ao tópico k no corpo, respectivamente (NEWMAN et al., [s.d.]).

Figura 3 - Modelo gráfico para LDA(*)



Fonte: (NEWMAN et al., [s.d.]).

(*)As variáveis observáveis (palavras) estão em cinza, os hiperparâmetros são mostrados em quadrados.

O primeiro teste realizado para verificar se todo o fluxo funcionava foi a extração de tópicos dos títulos dos documentos, pois é um volume menor de palavras para a avaliação. O fluxo funcionou, mas este teste inicial mostrou a necessidade de refinamentos serem feitos para garantir a qualidade dos tópicos.

Quadro 2: Teste de extração de tópicos dos títulos.

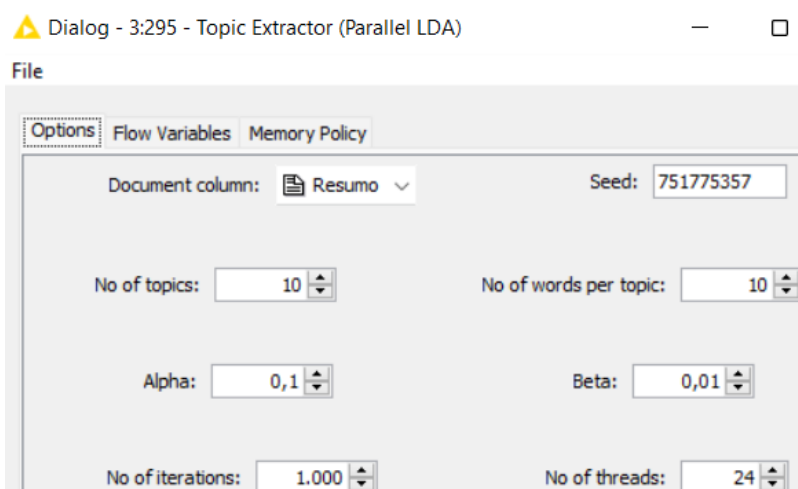
Palavras	1905 - 1999	2000 - 2010	2011 - 2021	1905 - 2021
Tópico 0	<i>des food les</i>	<i>using analysis based</i>	<i>effects food induced</i>	<i>production acid bacteria</i>
Tópico 1	<i>production acid food</i>	<i>del para alimentos</i>	<i>production acid bacteria</i>	<i>antioxidant oil activity</i>
Tópico 2	<i>foods detection food</i>	<i>food effects growth</i>	<i>del para des</i>	<i>food growth effects</i>
Tópico 3	<i>food industry development</i>	<i>antioxidant activity oil</i>	<i>properties effect quality</i>	<i>food using based</i>
Tópico 4	<i>food development effects</i>	<i>listeria food monocytogenes</i>	<i>based using detection</i>	<i>food safety industry</i>
Tópico 5	<i>food rats effects</i>	<i>food safety production</i>	<i>antioxidant activity oil</i>	<i>effect quality properties</i>
Tópico 6	<i>der und von</i>	<i>properties effect protein</i>	<i>food production water</i>	<i>listeria food detection</i>
Tópico 7	<i>food drying heat</i>	<i>food using drying</i>	<i>food review applications</i>	<i>food effects induced</i>
Tópico 8	<i>protein effect properties</i>	<i>determination food chromatography</i>	<i>determination food analysis</i>	<i>determination food detection</i>
Tópico 9	<i>determination food</i>	<i>food effects</i>	<i>food study</i>	<i>food production</i>

	<i>foods</i>	<i>dietary</i>	<i>safety</i>	<i>water</i>
Tópico 10	<i>food</i> <i>processing</i> <i>industry</i>	<i>quality</i> <i>effect</i> <i>storage</i>	<i>food</i> <i>production</i> <i>waste</i>	<i>properties</i> <i>wheat</i> <i>effect</i>
Tópico 11	<i>food</i> <i>production</i> <i>growth</i>	<i>production</i> <i>acid</i> <i>food</i>	<i>properties</i> <i>based</i> <i>food</i>	<i>food</i> <i>review</i> <i>applications</i>

O parâmetro alfa define a prioridade de Dirichlet nas distribuições de tópicos por documento e este alfa é utilizado para todos os tópicos. Ele define o peso prévio do tópico k em um documento. Normalmente, um número menor que 1 é usado quando se deseja poucos tópicos por documento e por isso na configuração padrão o alfa é 0,1, pois é comum esse desejo.

O parâmetro beta define a distribuição multinomial anterior por tópico sobre as palavras e este beta é utilizado para todas as palavras. Ele define o peso anterior da palavra w em um tópico. Normalmente um número muito menor que 1 é usado quando se deseja poucas palavras por tópico e por isso na configuração padrão o beta é 0,01, por ser um desejo comum.

Figura 4: Configurações padrão do nó *Topic Extractor (Parallel LDA)*.



Fonte: KNIME Analytics Platform

Levando em consideração o tempo de execução, escolheu-se realizar os testes de parâmetros com 500 iterações. Entretanto, a extração de tópicos final considerou 1.500 iterações. Os testes foram realizados com os resumos do período 1905 - 1999, por ser o período com menor quantidade de documentos e, conseqüentemente, com o maior risco. Para facilitar as diversas comparações, foram extraídos quatro tópicos em cada teste.

Quadro 3: Comparação tópicos para variações dos parâmetros

	Beta			
Alfa	0,0025	0,005	0,01	0,02
0,025	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, foods, products, intake. 2: food, production, products, industry. 3: food, temperature, products, method. 4: food, production, species, growth.	1: food, foods, products, intake. 2: food, production, products, industry. 3: food, temperature, products, method. 4: food, production, species, growth.
0,05	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, species, growth.	1: food, diet, dietary, intake. 2: food, production, products, industry. 3: food, temperature, products, acid. 4: food, production, species, growth.

0,1	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, foods, products, rats. 2: food, production, products, industry. 3: food, temperature, products, method. 4: food, production, species, growth.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, species, growth.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, species, growth.
0,2	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, growth, species.	1: food, production, products, industry. 2: food, temperature, water, process. 3: food, acid, products, protein. 4: food, production, species, growth.	1: food, foods, products, acid. 2: food, production, products, industry. 3: food, temperature, products, method. 4: food, production, species, growth.

Os mesmos tópicos apareceram para diferentes combinações, o que indica que precisam de alterações mais bruscas para interferir nos resultados. Como não houve uma combinação diferente do padrão que gerou um resultado considerado satisfatório, escolheu-se manter a condição padrão para a extração final de tópicos.

Além dos parâmetros, foram testadas a quantidade de palavras por tópico, uma vez que palavras a menos pode dificultar o entendimento do que é tal tópico, e palavras em excesso pode significar que certas palavras não se enquadram bem com as outras palavras, deixando o tópico confuso. Outro caso é quando palavras a mais não agregam ao sentido do tópico. Os testes foram realizados com o período todo

(1905 - 2021) por ele representar todos os recortes. Considerou-se três palavras como o melhor cenário.

Quadro 4: Teste quantidades de palavras por tópico.

Tópico	2 palavras	3 palavras	4 palavras	5 palavras
0	<i>production</i> <i>waste</i>	<i>production</i> <i>waste</i> <i>process</i>	<i>production</i> <i>waste</i> <i>process</i> <i>food</i>	<i>production</i> <i>waste</i> <i>process</i> <i>food</i> <i>energy</i>
1	<i>starch</i> <i>flour</i>	<i>starch</i> <i>flour</i> <i>content</i>	<i>starch</i> <i>flour</i> <i>content</i> <i>wheat</i>	<i>starch</i> <i>flour</i> <i>content</i> <i>wheat</i> <i>protein</i>
2	<i>temperature</i> <i>drying</i>	<i>temperature</i> <i>drying</i> <i>model</i>	<i>temperature</i> <i>drying</i> <i>model</i> <i>process</i>	<i>temperature</i> <i>drying</i> <i>model</i> <i>process</i> <i>heat</i>
3	<i>cells</i> <i>food</i>	<i>cells</i> <i>food</i> <i>effects</i>	<i>cells</i> <i>food</i> <i>effects</i> <i>cell</i>	<i>cells</i> <i>food</i> <i>effects</i> <i>cell</i> <i>induced</i>
4	<i>food</i> <i>production</i>	<i>food</i> <i>production</i> <i>agricultural</i>	<i>food</i> <i>production</i> <i>agricultural</i> <i>water</i>	<i>food</i> <i>production</i> <i>agricultural</i> <i>water</i> <i>agriculture</i>

	<i>data</i> 5 <i>model</i>	<i>data</i> <i>model</i> system	<i>data</i> <i>model</i> <i>system</i> based	<i>data</i> <i>model</i> <i>system</i> <i>based</i> <i>using</i>
	<i>food</i> 6 <i>products</i>	<i>food</i> <i>products</i> review	<i>food</i> <i>products</i> <i>review</i> foods	<i>food</i> <i>products</i> <i>review</i> <i>foods</i> <i>applications</i>
	<i>plant</i> 7 <i>soil</i>	<i>plant</i> <i>soil</i> crop	<i>plant</i> <i>soil</i> <i>crop</i> plants	<i>plant</i> <i>soil</i> <i>crop</i> <i>plants</i> yield
	<i>food</i> 8 <i>species</i>	<i>food</i> <i>species</i> fish	<i>food</i> <i>species</i> <i>fish</i> production	<i>food</i> <i>species</i> <i>fish</i> <i>production</i> <i>growth</i>
	<i>food</i> 9 <i>products</i>	<i>food</i> <i>products</i> exposure	<i>food</i> <i>products</i> <i>exposure</i> samples	<i>food</i> <i>products</i> <i>exposure</i> <i>samples</i> levels
	<i>species</i> 10 <i>fruit</i>	<i>species</i> <i>fruit</i> food	<i>species</i> <i>fruit</i> <i>food</i> plants	<i>species</i> <i>fruit</i> <i>food</i> <i>plants</i> plant

				<i>food products consumers study</i>
11	<i>food products</i>	<i>food products consumers</i>	<i>food products consumers</i> study	<i>food products consumers</i> consumer
12	<i>protein properties</i>	<i>protein properties</i> food	<i>protein properties food</i> water	<i>protein properties food water</i> stability
13	<i>antioxidant oil</i>	<i>antioxidant oil</i> activity	<i>antioxidant oil activity</i> acid	<i>antioxidant oil activity acid</i> compounds
14	<i>activity enzyme</i>	<i>activity enzyme</i> acid	<i>activity enzyme acid</i> production	<i>activity enzyme acid production</i> protein
15	<i>milk sensory</i>	<i>milk sensory</i> cheese	<i>milk sensory cheese</i> honey	<i>milk sensory cheese honey</i> products
16	<i>packaging food</i>	<i>packaging food</i> films	<i>packaging food films</i> materials	<i>packaging food films materials</i> properties

			<i>method</i> <i>detection</i> <i>food</i> <i>samples</i>	<i>method</i> <i>detection</i> <i>food</i> <i>analysis</i>
17	<i>method</i> <i>detection</i>	<i>method</i> <i>detection</i> <i>food</i>	<i>method</i> <i>detection</i> <i>food</i> <i>samples</i>	<i>method</i> <i>detection</i> <i>food</i> <i>analysis</i>
			<i>storage</i> <i>meat</i> <i>quality</i> <i>treatment</i>	<i>storage</i> <i>meat</i> <i>quality</i> <i>treatment</i> <i>processing</i>
18	<i>storage</i> <i>meat</i>	<i>storage</i> <i>meat</i> <i>quality</i>	<i>storage</i> <i>meat</i> <i>quality</i> <i>treatment</i>	<i>storage</i> <i>meat</i> <i>quality</i> <i>treatment</i> <i>processing</i>
			<i>food</i> <i>foods</i> <i>dietary</i> <i>intake</i>	<i>food</i> <i>foods</i> <i>dietary</i> <i>intake</i> <i>diet</i>
19	<i>food</i> <i>foods</i>	<i>food</i> <i>foods</i> <i>dietary</i>	<i>food</i> <i>foods</i> <i>dietary</i> <i>intake</i>	<i>food</i> <i>foods</i> <i>dietary</i> <i>intake</i> <i>diet</i>
			<i>food</i> <i>safety</i> <i>research</i> <i>products</i>	<i>food</i> <i>safety</i> <i>research</i> <i>products</i> <i>industry</i>
20	<i>food</i> <i>safety</i>	<i>food</i> <i>safety</i> <i>research</i>	<i>food</i> <i>safety</i> <i>research</i> <i>products</i>	<i>food</i> <i>safety</i> <i>research</i> <i>products</i> <i>industry</i>
			<i>bacteria</i> <i>strains</i> <i>food</i> <i>coli</i>	<i>bacteria</i> <i>strains</i> <i>food</i> <i>coli</i> <i>growth</i>
21	<i>bacteria</i> <i>strains</i>	<i>bacteria</i> <i>strains</i> <i>food</i>	<i>bacteria</i> <i>strains</i> <i>food</i> <i>coli</i>	<i>bacteria</i> <i>strains</i> <i>food</i> <i>coli</i> <i>growth</i>

Foram testados números de tópicos diferentes com todo o período. Poucos tópicos podem originar tópicos genéricos, e muitos tópicos podem ocasionar repetição de temáticas ou o aparecimento de uma temática pouco frequente. Descartou-se o total de 42 tópicos, pois certas palavras apareceram em mais de um tópico, parecendo uma repetição de temática. Além disso, alguns tópicos apresentaram variações da mesma palavra. Considerou-se, ainda, 12 tópicos como um número insuficiente para

o período inteiro, pois surgiram tópicos genéricos. 32 e 22 tópicos foram considerados bons o suficiente para o período inteiro. Para facilitar a comparação, foi escolhido extrair 22 tópicos.

Quadro 5: Teste de quantidade de tópicos (Período inteiro).

Tópico	2 palavras	3 palavras	4 palavras	5 palavras
0	<i>production</i> <i>waste</i>	<i>production</i> <i>waste</i> <i>process</i>	<i>production</i> <i>waste</i> <i>process</i> <i>food</i>	<i>production</i> <i>waste</i> <i>process</i> <i>food</i> <i>energy</i>
1	<i>starch</i> <i>flour</i>	<i>starch</i> <i>flour</i> <i>content</i>	<i>starch</i> <i>flour</i> <i>content</i> <i>wheat</i>	<i>starch</i> <i>flour</i> <i>content</i> <i>wheat</i> <i>protein</i>
2	<i>temperature</i> <i>drying</i>	<i>temperature</i> <i>drying</i> <i>model</i>	<i>temperature</i> <i>drying</i> <i>model</i> <i>process</i>	<i>temperature</i> <i>drying</i> <i>model</i> <i>process</i> <i>heat</i>
3	<i>cells</i> <i>food</i>	<i>cells</i> <i>food</i> <i>effects</i>	<i>cells</i> <i>food</i> <i>effects</i> <i>cell</i>	<i>cells</i> <i>food</i> <i>effects</i> <i>cell</i> <i>induced</i>
4	<i>food</i> <i>production</i>	<i>food</i> <i>production</i> <i>agricultural</i>	<i>food</i> <i>production</i> <i>agricultural</i> <i>water</i>	<i>food</i> <i>production</i> <i>agricultural</i> <i>water</i> <i>agriculture</i>

	<i>data</i> 5 <i>model</i>	<i>data</i> <i>model</i> system	<i>data</i> <i>model</i> <i>system</i> based	<i>data</i> <i>model</i> <i>system</i> <i>based</i> <i>using</i>
	<i>food</i> 6 <i>products</i>	<i>food</i> <i>products</i> review	<i>food</i> <i>products</i> <i>review</i> foods	<i>food</i> <i>products</i> <i>review</i> <i>foods</i> <i>applications</i>
	<i>plant</i> 7 <i>soil</i>	<i>plant</i> <i>soil</i> crop	<i>plant</i> <i>soil</i> <i>crop</i> plants	<i>plant</i> <i>soil</i> <i>crop</i> <i>plants</i> yield
	<i>food</i> 8 <i>species</i>	<i>food</i> <i>species</i> fish	<i>food</i> <i>species</i> <i>fish</i> production	<i>food</i> <i>species</i> <i>fish</i> <i>production</i> <i>growth</i>
	<i>food</i> 9 <i>products</i>	<i>food</i> <i>products</i> exposure	<i>food</i> <i>products</i> <i>exposure</i> samples	<i>food</i> <i>products</i> <i>exposure</i> <i>samples</i> levels
	<i>species</i> 10 <i>fruit</i>	<i>species</i> <i>fruit</i> food	<i>species</i> <i>fruit</i> <i>food</i> plants	<i>species</i> <i>fruit</i> <i>food</i> <i>plants</i> plant

				<i>food products consumers study</i>
11	<i>food products</i>	<i>food products consumers</i>	<i>food products consumers</i> study	<i>food products consumers</i> consumer
12	<i>protein properties</i>	<i>protein properties</i> food	<i>protein properties food</i> water	<i>protein properties food water</i> stability
13	<i>antioxidant oil</i>	<i>antioxidant oil</i> activity	<i>antioxidant oil activity</i> acid	<i>antioxidant oil activity acid</i> compounds
14	<i>activity enzyme</i>	<i>activity enzyme</i> acid	<i>activity enzyme acid</i> production	<i>activity enzyme acid production</i> protein
15	<i>milk sensory</i>	<i>milk sensory</i> cheese	<i>milk sensory cheese</i> honey	<i>milk sensory cheese honey</i> products
16	<i>packaging food</i>	<i>packaging food</i> films	<i>packaging food films</i> materials	<i>packaging food films materials</i> properties

				<i>method</i>
17	<i>method</i> <i>detection</i>	<i>method</i> <i>detection</i> food	<i>method</i> <i>detection</i> <i>food</i> samples	<i>detection</i> <i>food</i> <i>samples</i> analysis
18	<i>storage</i> <i>meat</i>	<i>storage</i> <i>meat</i> quality	<i>storage</i> <i>meat</i> <i>quality</i> treatment	<i>storage</i> <i>meat</i> <i>quality</i> <i>treatment</i> processing
19	<i>food</i> <i>foods</i>	<i>food</i> <i>foods</i> dietary	<i>food</i> <i>foods</i> <i>dietary</i> intake	<i>food</i> <i>foods</i> <i>dietary</i> <i>intake</i> diet
20	<i>food</i> <i>safety</i>	<i>food</i> <i>safety</i> research	<i>food</i> <i>safety</i> <i>research</i> products	<i>food</i> <i>safety</i> <i>research</i> <i>products</i> industry
21	<i>bacteria</i> <i>strains</i>	<i>bacteria</i> <i>strains</i> food	<i>bacteria</i> <i>strains</i> <i>food</i> coli	<i>bacteria</i> <i>strains</i> <i>food</i> <i>coli</i> growth

Quadro 6: Teste de quantidade de tópicos (1905 - 1999).

Tópico	12 tópicos	22 tópicos
0	<i>food</i> <i>production</i> <i>agricultural</i>	<i>oil</i> <i>acid</i> <i>fatty</i>

1	<i>food foods dietary</i>	<i>food irradiation safety</i>
2	<i>food packaging processing</i>	<i>content flour wheat</i>
3	<i>food species production</i>	<i>food dna cells</i>
4	<i>protein oil content</i>	<i>strains bacteria growth</i>
5	<i>food strains bacteria</i>	<i>food species development</i>
6	<i>waste der treatment</i>	<i>food weight diet</i>
7	<i>temperature model drying</i>	<i>system control food</i>
8	<i>food products industry</i>	<i>properties flow starch</i>
9	<i>food rats intake</i>	<i>der und die</i>
10	<i>acid protein activity</i>	<i>food rats effects</i>

11	<i>method food samples</i>	<i>food production fish</i>
12		<i>packaging food materials</i>
13		<i>meat fish beef</i>
14		<i>production treatment waste</i>
15		<i>protein milk proteins</i>
16		<i>food products foods</i>
17		<i>method food samples</i>
18		<i>production food soil</i>
19		<i>temperature model drying</i>
20		<i>food industry research</i>

21	<i>food</i> <i>foods</i> <i>dietary</i>
-----------	---

Quadro 7: Teste de quantidade de tópicos (2000 - 2010).

Tópico	12 tópicos	22 tópicos
0	<i>food</i> <i>intake</i> <i>effects</i>	<i>food</i> <i>intake</i> <i>foods</i>
1	<i>model</i> <i>using</i> <i>data</i>	<i>antioxidant</i> <i>activity</i> <i>extracts</i>
2	<i>food</i> <i>coli</i> <i>monocytogenes</i>	<i>packaging</i> <i>food</i> <i>materials</i>
3	<i>protein</i> <i>properties</i> <i>content</i>	<i>storage</i> <i>meat</i> <i>quality</i>
4	<i>acid</i> <i>production</i> <i>activity</i>	<i>temperature</i> <i>drying</i> <i>pressure</i>
5	<i>food</i> <i>packaging</i> <i>process</i>	<i>method</i> <i>samples</i> <i>food</i>
6	<i>oil</i> <i>acid</i> <i>antioxidant</i>	<i>bacteria</i> <i>growth</i> <i>strains</i>
7	<i>storage</i> <i>quality</i> <i>temperature</i>	<i>properties</i> <i>protein</i> <i>water</i>

8	<i>production food soil</i>	<i>food safety risk</i>
9	<i>food products foods</i>	<i>production soil crop</i>
10	<i>food species fish</i>	<i>activity enzyme acid</i>
11	<i>method samples food</i>	<i>food production research</i>
12		<i>starch flour wheat</i>
13		<i>pcr food dna</i>
14		<i>model data using</i>
15		<i>food foods products</i>
16		<i>food species fish</i>
17		<i>food cells effects</i>

18		<i>food allergy detection</i>
19		<i>oil fatty acid</i>
20		<i>production waste process</i>
21		<i>milk cheese sensory</i>

Quadro 8: Teste de quantidade de tópicos (2011 - 2021).

Tópico	12 tópicos	22 tópicos
0	<i>food intake effects</i>	<i>production waste acid</i>
1	<i>detection method samples</i>	<i>meat storage quality</i>
2	<i>food products review</i>	<i>food foods dietary</i>
3	<i>food study products</i>	<i>starch flour bread</i>
4	<i>production water soil</i>	<i>detection nanoparticles based</i>

5	<i>food bacteria strains</i>	<i>fish food species</i>
6	<i>food species fish</i>	<i>cells effects cell</i>
7	<i>antioxidant activity compounds</i>	<i>method samples honey</i>
8	<i>content products quality</i>	<i>fruit content fruits</i>
9	<i>properties food water</i>	<i>food review products</i>
10	<i>production acid fermentation</i>	<i>food safety risk</i>
11	<i>food process model</i>	<i>plant plants genetic</i>
12		<i>antioxidant oil activity</i>
13		<i>properties food films</i>
14		<i>food system based</i>

15		<i>detection food analysis</i>
16		<i>food consumers study</i>
17		<i>protein proteins activity</i>
18		<i>bacteria food strains</i>
19		<i>drying temperature process</i>
20		<i>food production agricultural</i>
21		<i>soil water production</i>

Quadro 9 - Teste de quantidade de tópicos de afiliações brasileiras.

Tópico	6 tópicos	12 tópicos
0	<i>food species production</i>	<i>drying temperature processing</i>
1	<i>compounds content antioxidant</i>	<i>films packaging food</i>

2	<i>food using method</i>	<i>samples method analysis</i>
3	<i>food strains bacteria</i>	<i>products flour sensory</i>
4	<i>food production oil</i>	<i>production using process</i>
5	<i>food production products</i>	<i>compounds antioxidant activity</i>
6		<i>food diet weight</i>
7		<i>food brazil study</i>
8		<i>food products review</i>
9		<i>strains food bacteria</i>
10		<i>oil properties fatty</i>
11		<i>production species soil</i>

2.3.5 Avaliação

As etapas realizadas foram revisadas e o modelo foi avaliado de forma mais aprofundada. Cada nó de todos os fluxos foi revisado, de modo a conferir se houve algum tipo de erro de configuração.

2.3.6 Implementação

O propósito deste modelo é aumentar o conhecimento. Portanto, a etapa de implementação neste contexto consistiu em organizar e apresentar o conhecimento de forma que outras pessoas possam utilizá-lo, como foi feito neste documento.

3. RESULTADOS E DISCUSSÃO

3.1 ANÁLISE MACRO

3.1.1 Áreas do conhecimento

Antes dos resultados serem limitados a certas disciplinas (busca intermediária), a consulta retornava 410.420 documentos e a distribuição de documentos por disciplinas estava organizada de acordo com a Tabela 1.

Tabela 1: Resultados da busca intermediária por área do conhecimento.

Área do conhecimento	Quantidade de documentos
<i>Agricultural and Biological Sciences</i>	180682
<i>Engineering</i>	66066
<i>Biochemistry, Genetics and Molecular Biology</i>	64985
<i>Chemistry</i>	53472
<i>Chemical Engineering</i>	33802
<i>Immunology and Microbiology</i>	33670
<i>Environmental Science</i>	31417
<i>Medicine</i>	23975
<i>Pharmacology, Toxicology and Pharmaceutics</i>	19267
<i>Materials Science</i>	15860
<i>Computer Science</i>	14334
<i>Physics and Astronomy</i>	12908
<i>Social Sciences</i>	11484

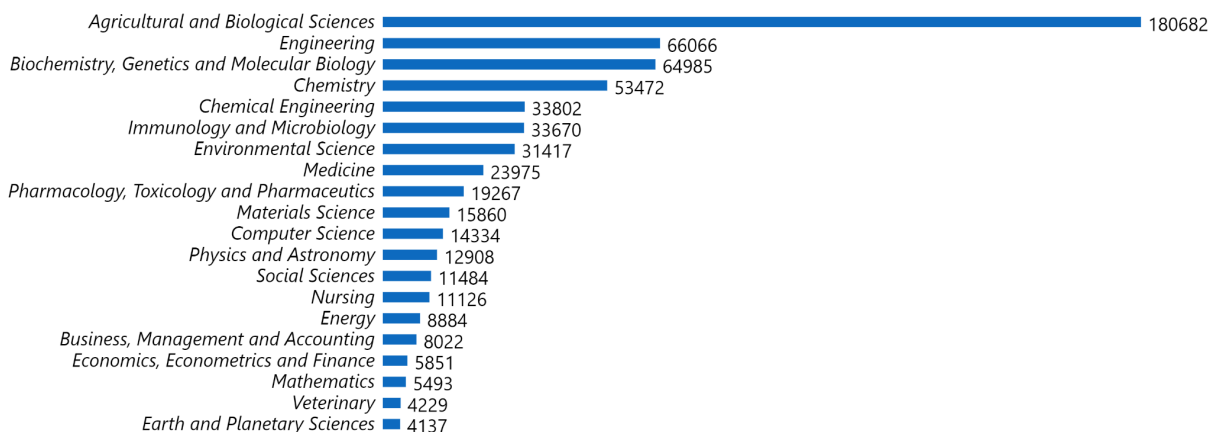
<i>Nursing</i>	11126
<i>Energy</i>	8884
<i>Business, Management and Accounting</i>	8022
<i>Economics, Econometrics and Finance</i>	5851
<i>Mathematics</i>	5493
<i>Veterinary</i>	4229
<i>Earth and Planetary Sciences</i>	4137
<i>Decision Sciences</i>	2672
<i>Health Professions</i>	2432
<i>Neuroscience</i>	2032
<i>Multidisciplinary</i>	1280
<i>Arts and Humanities</i>	1151
<i>Psychology</i>	412
<i>Dentistry</i>	97

Fonte: ("Scopus - Analyze search results | Busca intermediária", [s.d.])

O mesmo documento pode ser categorizado em mais de uma área, por isso o total de documentos é maior que a quantidade de resultados da busca.

Depois da escolha de áreas do conhecimento mais relacionadas com a temática da indústria de alimentos, o total de documentos retornados foi de 302.655. Na figura 5, é possível observar como essa escolha influenciou: *Medicine*, por exemplo, era a segunda área do conhecimento em número de documentos, mas depois se tornou a oitava.

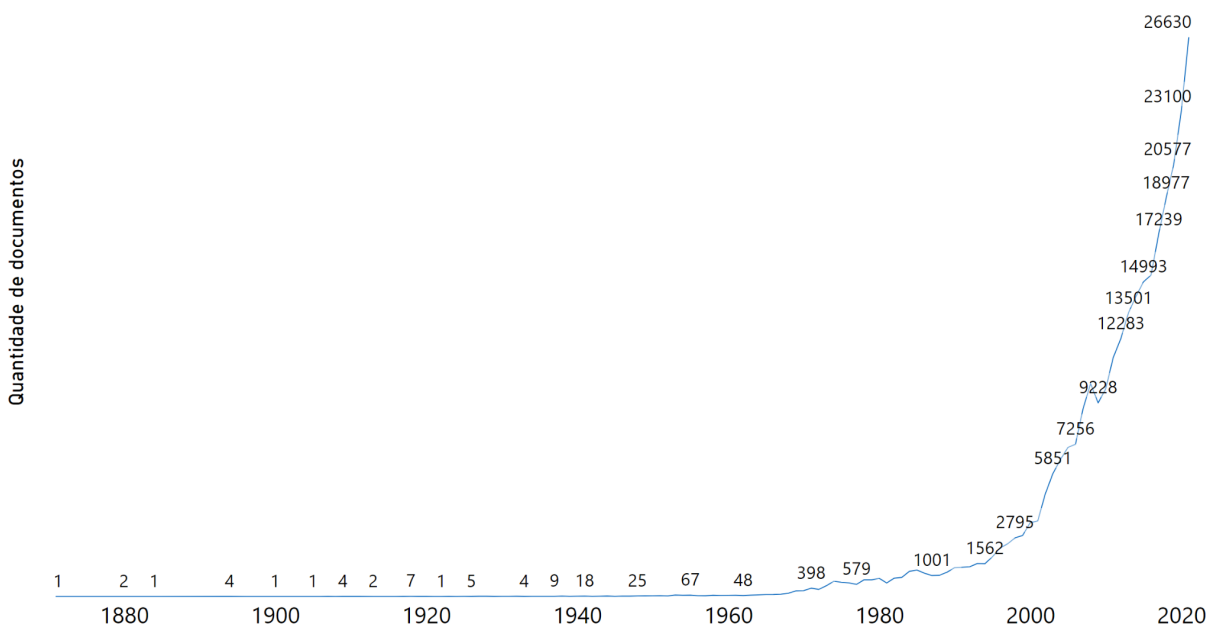
Figura 5: 20 áreas do conhecimento com mais documentos.



Dados: ("Scopus - Analyze search results | Busca final", [s.d.])

3.1.2 Evolução da quantidade nos anos

Figura 6: Resultados por ano de publicação.



Dados: ("Scopus - Analyze search results | Busca final", [s.d.])

A primeira publicação foi em 1871, mas até 1983 não passam de 1.000 documentos por ano. A Figura 6 mostra um crescimento exponencial, com um grande crescimento nas últimas décadas. Um ponto de destaque negativo foi 2009, quando houve uma queda de 9% em relação ao ano anterior e a diferença entre eles (913) é o maior decréscimo de um ano para o outro. Uma hipótese para esta queda é a influência da crise financeira de 2008, que pode ter reduzido investimentos em

pesquisa, especialmente no caso de novas investigações. Como pesquisas não geram resultados imediatamente, é esperado que a interferência não ocorra no mesmo ano. Um ponto curioso é que a pandemia (Covid-19) não provocou uma redução significativa na quantidade de documentos. 2021 foi o ano com maior crescimento no número absoluto de documentos com relação ao ano anterior: 3.530, o que representa um crescimento de 15,3%. O segundo maior crescimento em número absoluto foi em 2020, com 2.523 documentos a mais do que em 2019.

Neste estudo, foram considerados quatro recortes: 1905 -1999, 2000 - 2010, 2011 - 2021, 1905 - 2021. Foi escolhido 1905 ao invés de 1871 pois entre 1871 e 1904 os documentos estão sem resumo e foram desconsiderados na parte de extração de tópicos.

Apesar de 1905 -1999 ser um período longo, ele representa apenas 12% do total de documentos retornados pela busca, o que significa que os últimos 21 anos concentram 88% de todas as publicações utilizadas neste trabalho. 2000 - 2010 representou 25% do total, enquanto 2011 - 2021 é o período mais representativo com 62% de todos os documentos.

Tabela 2: Métricas descritivas dos 4 recortes de tempo.

	1905 -1999	2000 - 2010	2011 - 2021	1905 - 2021
Menor valor	0	3.514	11.388	0
Primeiro quartil	7	5.369	13.897	9
Mediana	35	7.104	15.313	72
Média	394	6.995	17.118	2.587
Terceiro quartil	646	9.086	19.777	1.418
Maior valor	2.908	10.141	26.630	26.630
Desvio padrão	666	2.383	4.762	5.349

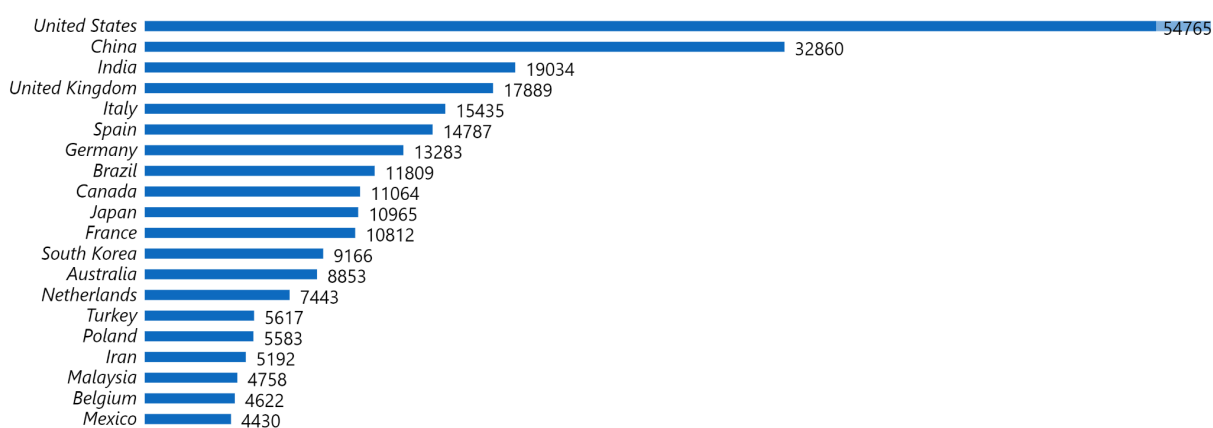
Dados: ("Scopus - Analyze search results | Busca final", [s.d.])

O crescimento na quantidade de documentos publicados pode também ser observado na Tabela 2. O maior valor do período 1905 - 1999 (2908 documentos em 1999) é menor que o menor valor do período 2000 - 2010 (3.514 documentos em 2000), o mesmo acontece quando comparamos o maior valor do período 2000 - 2010 (10.141 documentos em 2008) com o menor valor do período 2011 - 2021 (11.388 documentos em 2011). A média de documentos por ano entre 1905 a 1999 (394) é quase 18 vezes menor que a média do período 2000 - 2010 (6.995). Por outro lado, a média de documentos por ano do período 2011 - 2021(17.118) não chega a ser 3 vezes maior que a média de 2000 - 2010, o que significa que são períodos com uma diferença menor.

3.1.3 Documentos por local

Na Figura 7, é possível observar que os Estados Unidos é o local que produziu mais documentos (54.765), seguido pela China com 32.860 documentos, 21.905 documentos a menos que os Estados Unidos. O Brasil está em oitavo lugar no número de publicações, com 11.809 documentos.

Figura 7: 20 áreas com mais documentos.



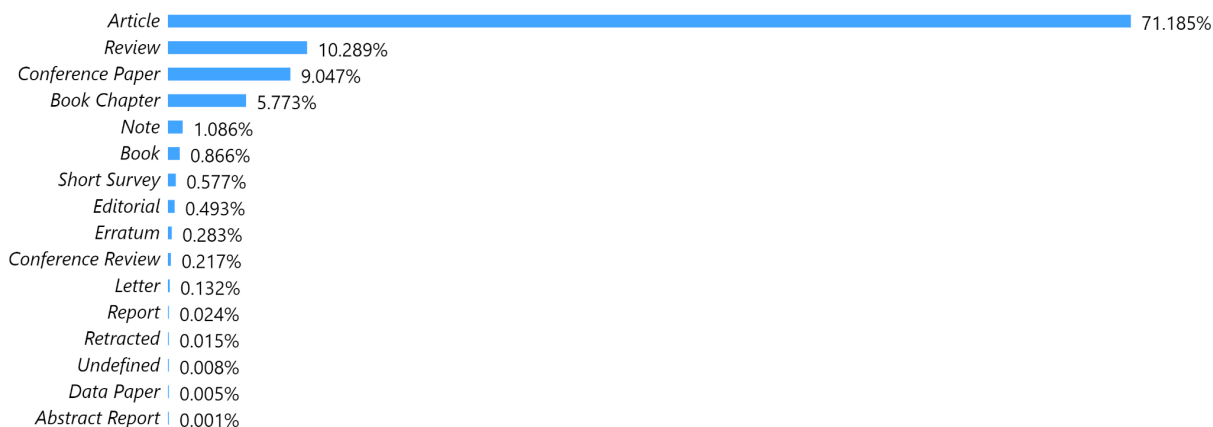
Dados: ("Scopus - Analyze search results | Busca final", [s.d.]).

3.1.4 Documentos por tipo

Na Figura 8, pode-se notar que o tipo de documento mais representativo é artigo, representando 71,2% dos documentos. Na etapa de extração de tópicos na qual apenas documentos com resumo foram considerados, o tipo artigo representou

73,9% dos documentos. Considerou-se que qualquer tipo de documento pode ser um indicativo de tendência, por isso nenhum foi excluído.

Figura 8: Proporção dos documentos por tipo.

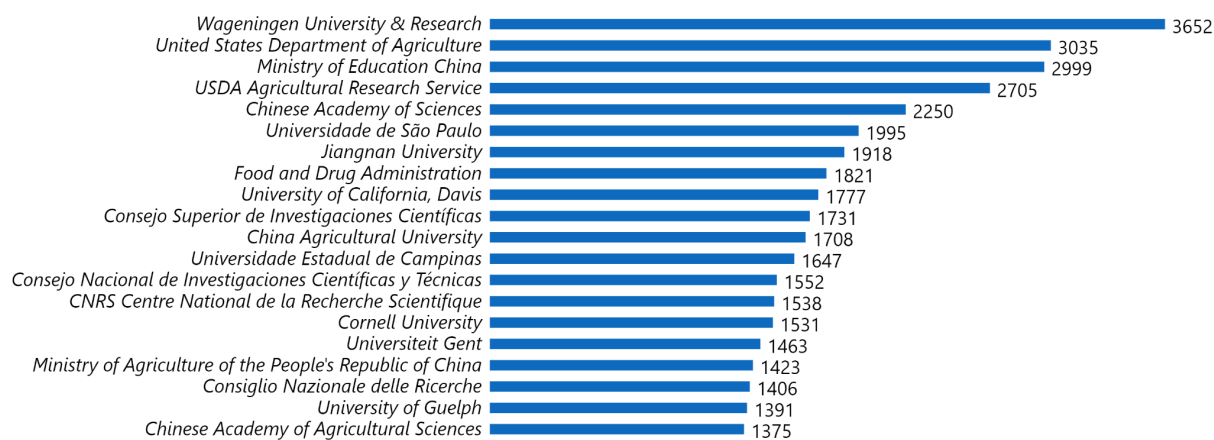


Dados: ("Scopus - Analyze search results | Busca final", [s.d.]).

3.1.5 Documentos por afiliação

Na figura 9, pode-se observar as instituições que produziram mais documentos. A instituição brasileira com maior número de documentos é a Universidade de São Paulo, com 1.995 documentos que correspondem ao sexto lugar. A Universidade Federal de Santa Catarina não está entre as 20 primeiras, ela possui 581 documentos associados a ela.

Figura 9: 20 instituições com mais documentos.

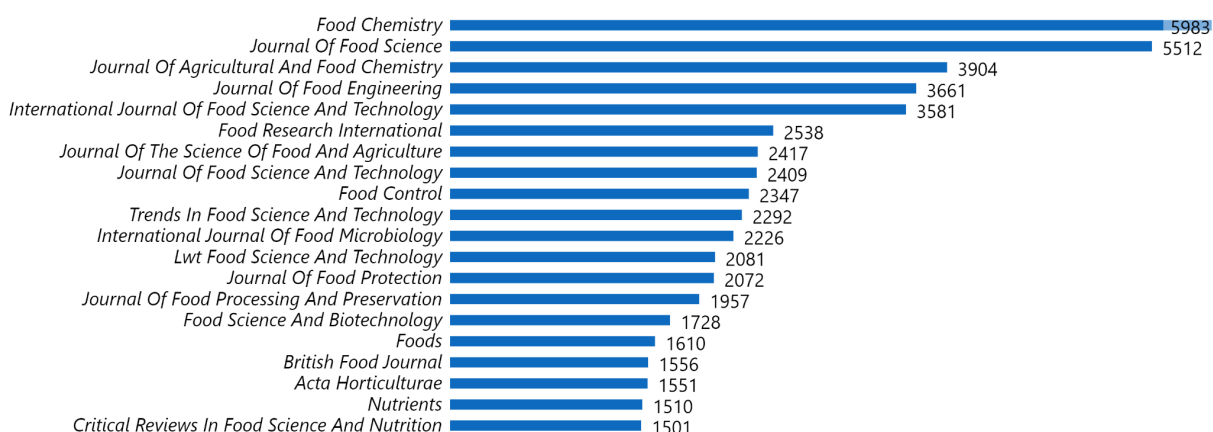


Dados: ("Scopus - Analyze search results | Busca final", [s.d.]).

3.1.6 Documentos por fonte

Na Figura 10, as principais fontes dos documentos podem ser observadas, um fator curioso é que a maioria dos nomes dela possuem a palavra *Food*, o que é um indicativo que a busca escolhida trouxe resultados coerentes.

Figura 10: 20 fontes com mais documentos.

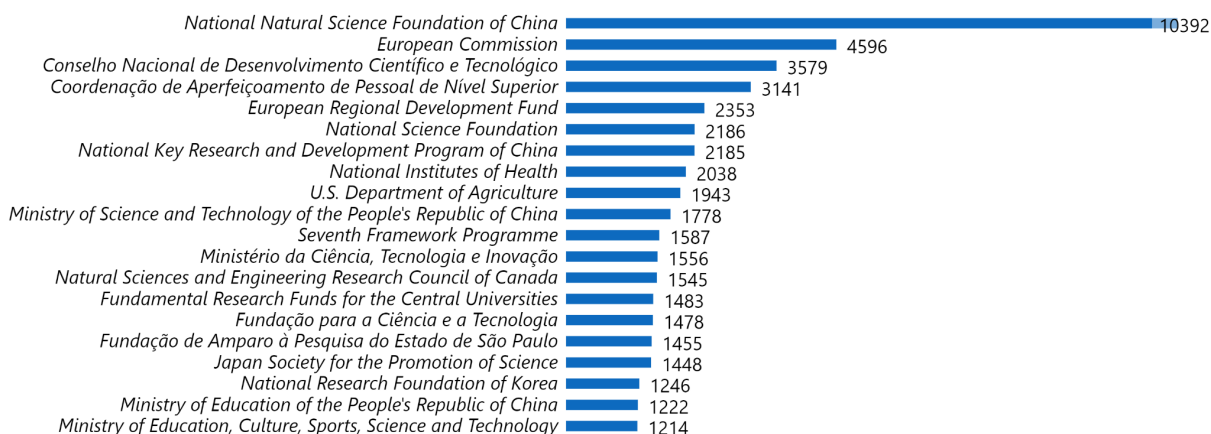


Dados: ("Scopus - Analyze search results | Busca final", [s.d.]).

3.1.7 Documentos por financiamento

Na Figura 11, os principais financiadores dessas publicações podem ser observados. Apesar do local com maior número de documentos ser os Estados Unidos, os maiores financiadores não são norte-americanos. O maior financiador é *National Natural Science Foundation of China*, seguido por *European Commission*. Vale mencionar que existem instituições brasileiras entre as maiores financiadoras.

Figura 11: 20 financiadores com mais documentos.

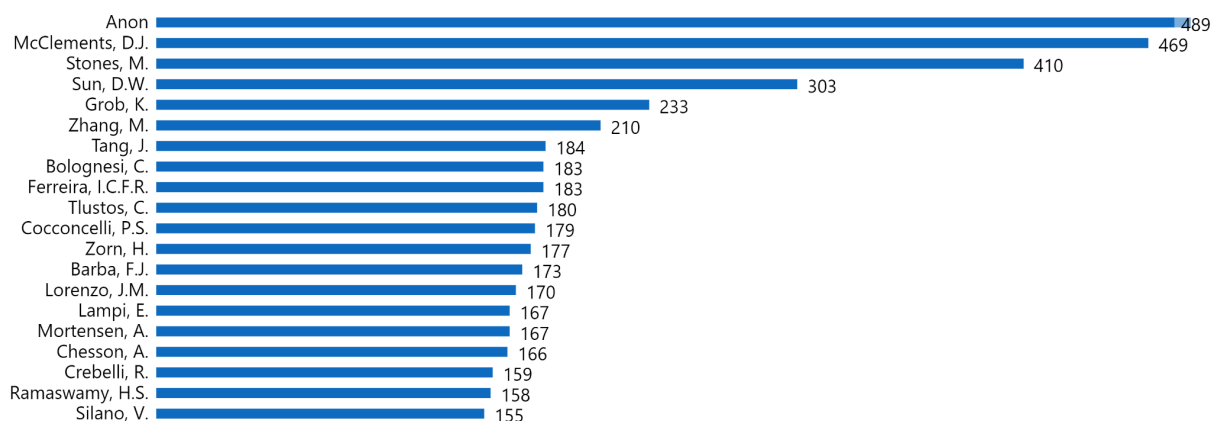


Dados: ("Scopus - Analyze search results | Busca final", [s.d.]

3.1.8 Documentos por autoria

Na Figura 12, pode-se observar que existem duas pessoas autoras que produziram mais de 400 documentos, o que é um número bem considerável.

Figura 12: 20 pessoas autoras com mais documentos.



Dados: ("Scopus - Analyze search results | Busca final", [s.d.]

Anon: sem autoria definida.

3.2 TÓPICOS

A partir dos tópicos extraídos durante a análise, termos genéricos foram descartados, assim como tópicos não estritamente relacionados com a engenharia de alimentos. Primeiro foram analisados os tópicos do período inteiro (1905 - 2021). O

tópico com mais documentos atribuídos é o tópico 2, que mostra como a secagem é um processo importante para a indústria. O segundo tópico em quantidade é a temática de óleos, que inclui atividade antioxidante. O terceiro é o tópico 21, que mostra como uma das maiores preocupações da indústria é o crescimento de bactérias. Essa observação é reforçada pelo quarto tópico em quantidade de documentos: tópico 20 *food safety*. O quinto tópico em quantidade trata da qualidade da carne no armazenamento (tópico 18).

É evidente que o período 1905 - 1999 trouxe como principal temática os sistemas de controle, pois o tópico 7 foi atribuído a mais documentos. A temática de secagem também apareceu neste período, mas como a segunda mais frequente (tópico 19). O terceiro tópico em quantidade também apareceu no período inteiro: crescimento bacteriano (tópico 4). Neste período, o conteúdo de farinha de trigo (tópico 3) também apareceu com relevância, sendo o quarto em quantidade. Dois tópicos foram elegíveis para o quinto lugar em quantidade: segurança e irradiação em alimentos (tópico 1) e materiais para embalagens de alimentos (tópico 12). Um ponto que chamou atenção foi o tópico 8, pois demonstra que, naquela época, ainda era comum a temática de propriedades fluidodinâmicas do amido.

O período 2000 - 2010, em comparação com os demais, teve o maior número de temáticas assertivas e não genéricas extraídas. Esse período também trouxe como principal temática o processo de secagem, mas com uma ênfase no emprego de variação da pressão durante a secagem (tópico 4). A segunda temática em maior quantidade versa sobre farinha de trigo e amido. A temática de crescimento de bactérias (tópico 6) também surgiu neste período, mas em terceiro lugar. O quarto tópico trata de *food safety* (tópico 8), mas trouxe um destaque para a palavra risco. O tópico de materiais para embalagens (tópico 2) também apareceu neste período, em quinto lugar. Um ponto que chamou a atenção é que a temática de alergia (tópico 18) surgiu neste período e, levando em conta a Resolução ANVISA/DC Nº 26 de 02/07/2015, ainda é um assunto bem atual.

O período 2011 - 2021 trouxe como temática principal secagem (tópico 19), o que também é um indicativo de como esse período possui um peso maior nos tópicos do período completo graças à quantidade de documentos desses últimos 10 anos. Em segundo em quantidade, observa-se a temática de crescimento bacteriano (tópico 18). Em terceiro lugar, tem-se o desperdício em produção de ácido (tópico 0). A temática de óleo e atividade antioxidante (tópico 12) também apareceu neste período,

ocupando o quarto lugar em quantidade de documentos atribuídos. A temática de filmes (tópico 13) apareceu em quinto lugar. Aqui entraram filmes indesejados (que surgem dentro de equipamentos), criação de filmes desejados (filme biodegradável por exemplo) e extração de propriedades de certos filmes. Um ponto que chamou a atenção é o surgimento da temática de nanopartículas (tópico 4), um assunto mais recente que ainda não era expressivo nos outros períodos. Esse aumento de relevância pode ter acontecido por causa da evolução tecnológica e dos processos relacionados.

Os tópicos que apareceram entre os cinco mais relevantes em todos os períodos foram: secagem e crescimento bacteriano. O tópico de secagem de alimentos possui muita sinergia com a situação da indústria de alimentos nas últimas décadas. Com a intensificação das exportações e importações de produtos alimentares, uma das principais preocupações da indústria é a conservação dos produtos, onde a secagem possui destaque. A secagem de alimentos ainda é um processo custoso para as empresas, e muitas pesquisas e documentos se propõem a trazer alternativas mais econômicas, sustentáveis e/ou com maior conservação de nutrientes. Já o tópico de crescimento bacteriano toma proporções globais à medida que novas normativas sanitárias são adotadas pelos países. O crescimento bacteriano é estudado na indústria de alimentos tanto para controle e entendimento de doenças alimentares, quanto para produtos produzidos através de bactérias. As palavras *food safety* apareceram em todos os períodos, assim como *production waste*.

Tabela 3: Quantidade de documentos atribuídos aos tópicos selecionados.

1905 - 2021		1905 - 1999		2000 - 2010		2011 - 2021	
Tópico	Quantidade	Tópico	Quantidade	Tópico	Quantidade	Tópico	Quantidade
Tópico 0: <i>production waste process</i>	12.607	Tópico 0: <i>oil acid fatty</i>	1.416	Tópico 1: <i>antiox- idant activity extracts</i>	2.736	Tópico 0: <i>production waste acid</i>	10.257
Tópico 1: <i>starch flour content</i>	13.664	Tópico 1: <i>food irradiation safety</i>	1.710	Tópico 2: <i>packaging food materials</i>	3.522	Tópico 1: <i>meat storage quality</i>	8.338

Tópico 2: <i>tempera- ture drying model</i>	18.900	Tópico 2: <i>content flour wheat</i>	1.855	Tópico 3: <i>storage meat quality</i>	3.014	Tópico 3: <i>starch flour bread</i>	9.378
Tópico 10: <i>species fruit food</i>	2.544	Tópico 3: <i>food dna cells</i>	974	Tópico 4: <i>tempera- ture drying pressure</i>	4.494	Tópico 4: <i>detection nanoparti- cles based</i>	6.402
Tópico 12: <i>protein properties food</i>	12.910	Tópico 4: <i>strains bacteria growth</i>	1.857	Tópico 6: <i>bacteria growth strains</i>	3.577	Tópico 8: <i>fruit content fruits</i>	5.296
Tópico 13: <i>antioxidant oil activity</i>	17.043	Tópico 7: <i>system control food</i>	3.376	Tópico 7: <i>properties protein water</i>	3.427	Tópico 10: <i>food safety risk</i>	6.829
Tópico 14: <i>activity enzyme acid</i>	12.157	Tópico 8: <i>properties flow starch</i>	1.569	Tópico 8: <i>food safety risk</i>	3.544	Tópico 12: <i>antioxidant oil activity</i>	10.073
Tópico 15: <i>milk sensory cheese</i>	8.184	Tópico 12: <i>packaging food materials</i>	1.710	Tópico 10: <i>activity enzyme acid</i>	2.825	Tópico 13: <i>properties food films</i>	9.971
Tópico 16: <i>packaging food films</i>	11.904	Tópico 13: <i>meat fish beef</i>	835	Tópico 12: <i>starch flour wheat</i>	3.585	Tópico 16: <i>food consumers study</i>	8.612
Tópico 18: <i>storage meat quality</i>	14.413	Tópico 14: <i>production treatment waste</i>	1.637	Tópico 13: <i>pcr food dna</i>	3.259	Tópico 17: <i>protein proteins activity</i>	6.607

Tópico 20: <i>food safety research</i>	16.359	Tópico 15: <i>protein milk proteins</i>	1.317	Tópico 18: <i>food allergy detection</i>	1.395	Tópico 18: <i>bacteria food strains</i>	10.555
Tópico 21: <i>bacteria strains food</i>	16.566	Tópico 19: <i>Temperature model drying</i>	3.040	Tópico 19: <i>oil fatty acid</i>	1.548	Tópico 19: <i>drying temperature process</i>	10.605
				Tópico 20: <i>production waste process</i>	2.755		
				Tópico 21: <i>milk cheese sensory</i>	2.662		

Escala de cor: variando entre azul para os valores mais baixos de tal coluna e rosa para os valores mais altos.

3.2.1 Tópicos de afiliações brasileiras

Apenas 11.809 documentos estão associados com afiliações brasileiras, 11.717 com *abstract* foram utilizados para extração de tópicos. Por representarem uma parcela pequena, não foram subdivididos em períodos e serão comparados com os tópicos do período total sem a restrição de país. Aqui a temática de secagem (tópico 0) também apareceu com o maior número de documentos atribuídos. O segundo em maior quantidade é a temática de farinha (tópico 3). A temática de atividade antioxidante (tópico 5) também aparece aqui, sendo o terceiro tópico em quantidade. Porém, deve-se observar que o tópico 10 pode estar relacionado e talvez o peso desse tema seja maior do que parece, especialmente por ele ser o quarto em quantidade.

Um ponto que chamou a atenção é que não surgiu uma temática similar à qualidade da carne (tópico 18 de todo o período) nas publicações brasileiras. Como o Brasil é um grande produtor de proteína animal, era esperado que tal tema aparecesse.

Tabela 4: Quantidade de documentos atribuídos aos tópicos selecionados de afiliações brasileiras.

Tópico	Palavras	Quantidade
0	<i>drying</i> <i>temperature</i> <i>processing</i>	1030
1	<i>films</i> <i>packaging</i> <i>food</i>	579
3	<i>products</i> <i>flour</i> <i>sensory</i>	975
5	<i>compounds</i> <i>antioxidant</i> <i>activity</i>	946
9	<i>strains</i> <i>food</i> <i>bacteria</i>	811
10	<i>oil</i> <i>properties</i> <i>fatty</i>	824

4. CONCLUSÃO

Através da realização da análise de dados disponíveis no banco Scopus usando a metodologia CRISP-DM e extração de tópicos com LDA, foi possível identificar os tópicos mais frequentes e através deles identificar tendências, sendo os objetivos do trabalho atingidos.

Com a extração de tópicos, foi possível perceber como as temáticas de secagem, segurança alimentar, atividade antioxidante e crescimento bacteriano são frequentes e, conseqüentemente, relevantes para a indústria de alimentos. Na análise macro, percebeu-se que 88% dos documentos foram publicados nos últimos 21 anos e que a quantidade de publicações está crescendo de forma exponencial, o que indica um futuro promissor.

Em trabalhos futuros, é recomendável utilizar mais de uma base de dados, de forma a tornar o estudo ainda mais representativo da realidade de publicações. Além disso, a utilização de dados de mercado seria de grande utilidade para identificar tendências fora do universo de pesquisa e tornar ainda mais abrangente o estudo. Outra possível nova abordagem é tornar a busca mais restritiva, para evitar resultados não relacionados com a engenharia de alimentos.

REFERÊNCIAS

AUGUSTO, P. E. Challenges, trends and opportunities in food processing. *Current Opinion in Food Science, Food Engineering & Processing*. v. 35, p. 72–78, 1 out. 2020.

AZEVEDO, A.; SANTOS, M. F. KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW. p. 6, [s.d.].

ELSEVIER. High-Quality Data - How Scopus Works - Scopus | Elsevier solutions. Disponível em: <<https://www.elsevier.com/solutions/scopus/how-scopus-works/high-quality-data>>. Acesso em: 12 fev. 2022.

HELDMAN, D. R.; LUND, D. B. The Beginning, Current, and Future of Food Engineering: A Perspective. *Food Engineering Series*, p. 3–18, 2011.

NEWMAN, D. et al. *Distributed Algorithms for Topic Models*. [s.d.].

Scopus - Analyze search results | Busca final. Disponível em: <<https://www.scopus.com/term/analyzer.uri?sid=c06dd4e890c183c53ba878285a5db6b5&origin=resultslist&src=s&s=TITLE-ABS-KEY+%28+%28food+w%2f10+engineer%29+OR+%28food+w%2f10+industry%29+OR+%28food+w%2f10+process%29+OR+%28food+w%2f10+tech%29+OR+%28food+w%2f10+science%29+OR+%28food+w%2f10+manufact%29+OR+%28food+w%2f10+product%29+OR+%28food+w%2f10+develop%29+OR+%28food+w%2f10+innovat%29+%29+AND+PUBYEAR+%3c+2022&sort=plf-f&sdt=a&sot=a&sl=247&count=302664&analyzeResults=Analyze+results&cluster=scosubjabbr%2c%22AGRI%22%2ct%2c%22ENGI%22%2ct%2c%22BIOC%22%2ct%2c%22CHEM%22%2ct%2c%22CENG%22%2ct%2c%22IMMU%22%2ct%2c%22PHAR%22%2ct%2c%22MATE%22%2ct%2c%22COMP%22%2ct%2c%22PHYS%22%2ct%2c%22MATH%22%2ct&txGid=e422e9008e6748268dcac853b76bf8d2>>. Acesso em: 11 mar. 2022.

Scopus - Analyze search results | Busca intermediária. Disponível em: <<https://www.scopus.com/term/analyzer.uri?sid=0f05a92a323aa482be7dd0587eae2df8&origin=resultslist&src=s&s=TITLE-ABS-KEY+%28+%28food+w%2f10+engineer%29+OR+%28food+w%2f10+industry%29+OR+%28food+w%2f10+process%29+OR+%28food+w%2f10+tech%29+OR+%28food+w%2f10+science%29+OR+%28food+w%2f10+manufact%29+OR+%28food+w%2f10+product%29+OR+%28food+w%2f10+develop%29+OR+%28food+w%2f10+innovat%29+%29+AND+PUBYEAR+%3c+2022&sort=plf-f&sdt=a&sot=a&sl=247&count=410420&analyzeResults=Analyze+results&txGid=9bf06f85450d5983d655e3d2eef4d59>>. Acesso em: 6 mar. 2022.

UYSAL, A. K.; GUNAL, S. The impact of preprocessing on text classification. *Information Processing & Management*, v. 50, n. 1, p. 104–112, 1 jan. 2014.

What does “Relevance” mean in Scopus? - Scopus: Access and use Support Center. Disponível em: <https://service.elsevier.com/app/answers/detail/a_id/14182/supporthub/scopus/kw/r>

elevance/>. Acesso em: 27 fev. 2022.

APÊNDICE

Quadro 12: resultados da pesquisa utilizada na introdução

Ano	Título	Ferramenta	Pesquisa	Ordenação
2022	<i>Advances in extrusion-dripping encapsulation of probiotics and omega-3 rich oils</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Health conscious consumers and sugar confectionery: Present aspects and projections</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Biofilm control strategies in food industry: Inhibition and utilization</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Characterisation of the volatile profile of microalgae and cyanobacteria using solid-phase microextraction followed by gas chromatography coupled to mass spectrometry</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>A systematic review of fermented Saccharina japonica: Fermentation conditions, metabolites, potential health benefits and mechanisms</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Improving Sustainable Environment of Biopolymers Using Nanotechnology</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Responsibility of the company's social impact and environmental orientation in marketing performance in health food companies</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>CRISPR/Cas12a-based technology: A powerful tool for biosensing in food safety</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Aloe vera: From ancient knowledge to the patent and innovation landscape—A review</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>The Phenomenon of Amelioration Usage as Food Business Development Strategy</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Applications of Metal Nanoparticles in the Agri-Food sector</i>	Google Scholar	"food industry" trends	Mais recente

2022	<i>Comparative Study of Corporate Sustainability in Proactive Environmental Management: The Case Study of Business Organization in Thailand</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>The Influence of Attitude and Need for Cognition on Student's Purchase Intention Behavior on Halal Food: Schools Clustering Perspective</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>A review of factors affecting the stability of zein-based nanoparticles loaded with bioactive compounds: from construction to application</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>News and Trends in the Development of Functional Foods: Probiotic Dairy and Non-Dairy Products</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Effects of konjac glucomannan on pasting, rheological, and structural properties of low-amylose rice starch</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>CONSUMPTION, PRODUCTION AND SAFETY OF FOOD AND BEVERAGES IN BULGARIA IN THE CONDITIONS OF COVID PANDEMIC</i> Vladi KURSHUMOV	Google Scholar	"food industry" trends	Mais recente
2022	<i>Recent Advances in the Food Application of Electrospun Nanofibers</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Isomaltulose: From origin to application and its beneficial properties-A bibliometric approach</i>	Google Scholar	"food industry" trends	Mais recente
2022	<i>Towards Sustainable Innovation in the Bakery Sector—An Example of Fibre-Enriched Bread</i>	Google Scholar	"food industry" trends	Mais recente
2018	<i>Emerging trends in European food, diets and food industry</i>	Google Scholar	"food industry" trends	Relevância
2013	<i>Innovation trends in the food industry: The case of functional foods</i>	Google Scholar	"food industry" trends	Relevância

2019	<i>Recent trends and applications of cellulose nanocrystals in food industry</i>	Google Scholar	"food industry" trends	Relevância
2017	<i>Current status and future trends of high-pressure processing in food industry</i>	Google Scholar	"food industry" trends	Relevância
2017	<i>A review on recent advances in cold plasma technology for the food industry: Current applications and future trends</i>	Google Scholar	"food industry" trends	Relevância
2009	<i>A review of wireless sensor technologies and applications in agriculture and food industry: state of the art and current trends</i>	Google Scholar	"food industry" trends	Relevância
2020	<i>Review of energy efficiency technologies in the food industry: trends, barriers, and opportunities</i>	Google Scholar	"food industry" trends	Relevância
2016	<i>Electrolyzed water as a novel sanitizer in the food industry: current trends and future perspectives</i>	Google Scholar	"food industry" trends	Relevância
2018	<i>Food industry digitalization: from challenges and trends to opportunities and solutions</i>	Google Scholar	"food industry" trends	Relevância
2017	<i>The agricultural production and food industry development trends in the context of food security of Russia</i>	Google Scholar	"food industry" trends	Relevância
2017	<i>New trends for a classical enzyme: Papain, a biotechnological success story in the food industry</i>	Google Scholar	"food industry" trends	Relevância
2021	<i>Garcinia brasiliensis fruits and its by-products: Antioxidant activity, health effects and future food industry trends—A bibliometric review</i>	Google Scholar	"food industry" trends	Relevância
2017	<i>New trends in the food industry: application of nanosensors in food packaging</i>	Google Scholar	"food industry" trends	Relevância
2009	<i>Use of enzymatic biosensors as quality indices: A synopsis of present and future trends in the food industry</i>	Google Scholar	"food industry" trends	Relevância
2013	<i>Trends in the acquisition of external</i>	Google	"food industry" trends	Relevância

	<i>knowledge for innovation in the food industry</i>	Scholar	trends	
2006	<i>Corporate social responsibility in the supply chain: an application in the food industry</i>	Google Scholar	"food industry" trends	Relevância
2020	<i>Trends in the food and sports nutrition industry: A review</i>	Google Scholar	"food industry" trends	Relevância
2012	<i>Trends in product recalls within the agri-food industry: Empirical evidence from the USA, UK and the Republic of Ireland</i>	Google Scholar	"food industry" trends	Relevância
2022	<i>Nanocellulose: Recent trends and applications in the food industry</i>	Google Scholar	"food industry" trends	Relevância
2008	<i>Sustainability in the agri-food industry: a literature review and overview of current trends</i>	Google Scholar	"food industry" trends	Relevância
2010	<i>Nanoencapsulation: a new trend in food engineering processing</i>	Google Scholar	"food engineering" trends	Relevância
2020	<i>Superwettability-based systems: Basic concepts, recent trends and future prospects for innovation in food engineering</i>	Google Scholar	"food engineering" trends	Relevância
2000	<i>Trends in food engineering</i>	Google Scholar	"food engineering" trends	Relevância
2008	<i>New food engineering research trends</i>	Google Scholar	"food engineering" trends	Relevância
2017	<i>Microwave Blanching: An Emerging Trend in Food Engineering and its Effects on Capsicum annum L</i>	Google Scholar	"food engineering" trends	Relevância
2012	<i>Trends in Vital Food and Control Engineering</i>	Google Scholar	"food engineering" trends	Relevância
2002	<i>Trends in food engineering</i>	Google Scholar	"food engineering" trends	Relevância

			trends	
2018	<i>Biosensors for sustainable food engineering: challenges and perspectives</i>	Google Scholar	"food engineering" trends	Relevância
2013	<i>Current trends in green technologies in food production and processing</i>	Google Scholar	"food engineering" trends	Relevância
2008	<i>Applications of microfluidic devices in food engineering</i>	Google Scholar	"food engineering" trends	Relevância
1997	<i>Handbook of food engineering practice</i>	Google Scholar	"food engineering" trends	Relevância
2015	<i>Trends in encapsulation technologies for delivery of food bioactive compounds</i>	Google Scholar	"food engineering" trends	Relevância
2019	<i>Introductory Chapter: A General Overview on Latest Trends in Food Engineering</i>	Google Scholar	"food engineering" trends	Relevância
2013	<i>Challenges facing food engineering</i>	Google Scholar	"food engineering" trends	Relevância
2021	<i>Electrospun nanofibers food packaging: Trends and applications in food systems</i>	Google Scholar	"food engineering" trends	Relevância
2001	<i>Introduction to food engineering</i>	Google Scholar	"food engineering" trends	Relevância
2011	<i>Modelling and analysis of complex food systems: state of the art and new trends</i>	Google Scholar	"food engineering" trends	Relevância
2002	<i>Unit operations in food engineering</i>	Google Scholar	"food engineering" trends	Relevância

			trends	
2020	Challenges, trends and opportunities in food processing	Google Scholar	"food engineering" trends	Relevância
2016	<i>Technological and engineering trends for production of gluten-free beers</i>	Google Scholar	"food engineering" trends	Relevância
2022	<i>Effects of konjac glucomannan on pasting, rheological, and structural properties of low-amylose rice starch</i>	Google Scholar	"food engineering" trends	Mais recente
2021	DRYING TECHNOLOGIES IN FOOD ENGINEERING	Google Scholar	"food engineering" trends	Mais recente
2022	<i>The effect of polysaccharide blends and salts on the glass transition temperature of the monosaccharide glucose</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Fractal Microstructure of Foods</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Nondestructive Methods for the Quality Assessment of Fruits and Vegetables Considering Their Physical and Biological Variability</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>Feeding the Future—Challenges and Limitations</i>	Google Scholar	"food engineering" trends	Mais recente
2021	ULTRASOUND ASSISTED EXTRACTION OF ANTIDIABETIC COMPOUNDS AND THEIR ENCAPSULATION	Google Scholar	"food engineering" trends	Mais recente
2021	A SYSTEMATIC REVIEW AND NETWORK ANALYSIS IN PRECISION DAIRY MONITORING	Google Scholar	"food engineering" trends	Mais recente

2021	<i>Designing delivery systems for functional ingredients by protein/polysaccharide interactions</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Strategies to mitigate protein deficit</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Recent developments in fermentation technology: toward the next revolution in food production</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Food engineering innovations across the food supply chain: debrief and learnings from the ICEF13 congress and the future of food engineering</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Inverse problems in food engineering: A review</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>Discussion on How to Improve the Teaching Efficacy of Chinese University Teachers from the Perspective of Information Leadership</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>Nonthermal Processing Technologies for Stabilization and Enhancement of Bioactive Compounds in Foods</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>APPLICATION OF NON-DESTRUCTIVE METHODS FOR SORTING OF RASPBERRIES</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Evaluation of maximum wall shear stress from air impingement to remove food deposits from stainless steel surfaces</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>The NOVA classification system: A critical perspective in food science</i>	Google Scholar	"food engineering" trends	Mais recente
2021	<i>Stabilnost minimalno procesiranog krumpira tretiranog fenolnim ekstraktom komorača</i>	Google Scholar	"food engineering" trends	Mais recente

2021	<i>Factors affecting contamination and infiltration of Escherichia coli K12 into spinach leaves during vacuum cooling</i>	Google Scholar	"food engineering" trends	Mais recente
2022	<i>Towards intelligent packaging: BCP-EVOH@optode for milk freshness measurement</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2022	<i>Advantages of nanotechnology developments in active food packaging</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2022	<i>Potato protein: current review of structure, technological properties, and potential application on spray drying microencapsulation</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2022	<i>Mannitol Production by Heterofermentative Lactic Acid Bacteria: a Review</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2022	<i>Application of blockchain technology in shaping the future of food industry based on transparency and consumer trust</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente

2022	<i>Membrane applications in the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Effects of different ingredients and stabilisers on properties of mixes based on almond drink for vegan ice cream production</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Rheological and pipe flow properties of chocolate masses at different temperatures</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Biochemical identification of a hyperthermostable L-ribulose 3-epimerase from <i>Labedella endophytica</i> and its application for D-allulose bioconversion</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Information needs assessment of decision making by stakeholders of the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Polyphenolic antibacterials for food preservation: Review, challenges, and current applications</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente

			w/3 trends))	
2021	<i>Applications of plant polymer-based solid foams: Current trends in the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Bioactive potential, health benefits and application trends of Syzygium malaccense (Malay apple): A bibliometric review</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Quality improvement of cakes with buckwheat flour, and its comparison with local branded cakes</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>Tubular cellulose from orange juice by-products as carrier of chemical preservatives; delivery kinetics and microbial stability of orange juice</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>How do the different types of carrier and drying techniques affect the changes in physico-chemical properties of powders from chokeberry pomace extracts?</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Mais recente
2021	<i>The influence of temperature, storage conditions, ph, and ionic strength on the</i>	Scopus	TITLE-ABS-KEY(("food	Mais recente

	<i>antioxidant activity and color parameters of rowan berry extracts</i>		<i>engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	
2021	<i>Technological properties of chickpea (Cicer arietinum): Production of snacks and health benefits related to type-2 diabetes</i>	Scopus	<i>TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	Mais recente
2021	<i>Current progress in the utilization of soy-based emulsifiers in food applications—a review</i>	Scopus	<i>TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	Mais recente
2021	<i>Microorganisms, the ultimate tool for clean label foods?</i>	Scopus	<i>TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	Mais recente
2000	<i>The 2nd annual study of purchasing trends among food industry buyers & specifiers</i>	Scopus	<i>TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	Relevância
1981	<i>Trends in Dairy Food Engineering</i>	Scopus	<i>TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i>	Relevância

1988	<i>Course development—Consumer trends in the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2016	<i>Six top trends to dominate food industry of 2025</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2006	<i>Employment trends in the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2013	<i>Seven top trends for the food industry in 2013: LFR</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2011	<i>Introduction to the special issue on recent trends in the food industry and food Chain</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
1975	<i>The influence of retailing trends on the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância

			w/3 trends))	
2006	<i>Section XI. Functional Foods: Functional Foods: International Considerations</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
1999	<i>Nutraceutical industry trends</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2015	<i>Latest trends in the pet food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2009	<i>Mixing in the Food Industry: Trends and Challenges</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2022	<i>Membrane applications in the food industry</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2017	<i>Introduction</i>	Scopus	TITLE-ABS-KEY(("food	Relevância

			engineering" w/3 trends) OR ("food industry" w/3 trends))	
2003	<i>Hot Topic Sessions Take on Industry's Hottest Issues</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2003	Future trends in food engineering	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
1984	QUALITY TRENDS IN THE FOOD INDUSTRY.	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
1984	QUALITY TRENDS IN THE FOOD INDUSTRY.	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância
2018	<i>Pea protein isolates: Structure, extraction, and functionality</i>	Scopus	TITLE-ABS-KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))	Relevância

1984	<p><i>QUALITY TRENDS IN THE FOOD INDUSTRY OF HUNGARY.</i></p>	<p><i>Scopus</i></p>	<p><i>TITLE-ABS- KEY(("food engineering" w/3 trends) OR ("food industry" w/3 trends))</i></p>	<p><i>Relevância</i></p>
------	---	----------------------	---	--------------------------