



Das patentes às disciplinas de sala de aula: uma abordagem de análise computacional

Flavio Izo, UFES/IFES, fizo@ifes.edu.br, <https://orcid.org/0000-0001-7189-0387>
Elias Oliveira, UFES, elias@lcad.inf.ufes.br, <https://orcid.org/0000-0003-2066-7980>

Resumo: Os planos de ensino orientam os estudos dos alunos, proporcionando a oportunidade de desenvolver projetos que podem resultar em patentes. Neste trabalho, apresentamos uma metodologia de análise computacional de dados visando conectar os conteúdos químicos e avaliar o quanto um curso de química aborda os temas patenteados. Esta análise permite identificar lacunas de conhecimento que os alunos precisam aprimorar para desenvolver ou melhorar uma patente específica. Foram utilizados 28 planos de ensino de um curso superior em Engenharia Química e 30 patentes químicas do INPI. Os resultados indicaram que, em média, metade dos conteúdos das patentes foram abordados ao final do terceiro período do curso, com Métrica Base (MB) e Métrica com Peso (MP) marcando 52% e 50%, respectivamente.

Palavras-chave: Inteligência Artificial, Recuperação da informação, Reconhecimento de Entidades Nomeadas, Plano de Ensino, Patente.

From Patents to Classroom Subjects: A Computational Analysis Approach

Abstract: The teaching plans guide students' studies, providing the opportunity to develop projects that may result in patents. In this study, we present a computational data analysis methodology aiming to link chemical contents and assess how much a chemistry course covers patented topics. This analysis helps identify knowledge gaps that students need to improve to develop or enhance a specific patent. We used 28 teaching plans from a Chemical Engineering undergraduate program and 30 chemical patents from INPI. Results indicated that, on average, half of the patent contents were covered by the end of the third semester, with Basic Metric (MB) and Weighted Metric (MP) scoring 52% and 50%, respectively.

Keywords: Artificial Intelligence, Information Retrieval, Named Entity Recognition, Teaching Plan, Patent.

1. Introdução

Os documentos científicos são fontes de conhecimento para diversas pesquisas e um desses documentos são as patentes. Esses documentos armazenam uma imensa fonte de conhecimento para comunidades de pesquisa e inovação em todo o mundo (KRESTEL *et al.*, 2021). Nos últimos anos, o número de publicações de patentes vêm aumentando e se tornando um enorme desafio para a análise e extração de informação de forma eficiente (KRESTEL *et al.*, 2021; GREENBERG; COHEN; GREWAL, 2021). Apesar dos desafios, as informações contidas nas patentes representam conhecimentos adquiridos pelos inventores por meio de estudos e pesquisas, tornando-se valiosas fontes de conhecimento tecnológico (KALTHAUS, 2020).

Outro documento importante são os planos de ensino dos componentes curriculares que estão presentes nos cursos de formação dos alunos. Eles desempenham um papel crucial ao orientar os estudos e destacar conceitos e habilidades necessárias para os alunos. Além disso, esses planos de ensino podem oferecer oportunidades para os alunos se envolverem em projetos de pesquisa que os capacitam a aplicar seu conhecimento de maneira prática. Desta forma, os alunos são incentivados a explorar áreas de interesse, identificar problemas e buscar soluções inovadoras. Por meio desse



processo, os alunos podem até mesmo desenvolver ideias originais que possam ser transformadas em patentes, ou aproveitar o conhecimento de uma patente e se desenvolver através dele (KALTHAUS, 2020).

As patentes químicas representam uma fonte valiosa de informações sobre novos conteúdos químicos He *et al.* (2021) e os últimos anos têm mostrado um aumento substancial dessas publicações. Isto levou a um interesse crescente na extração de informações (por exemplo, entidades nomeadas) destas publicações (SAAD, 2020). No entanto, dada a complexidade dos textos de patentes, a análise bibliométrica é mais comum na literatura, pois envolve a análise de dados organizados e estruturados, mais fáceis de processar do que dados não estruturados (PUC CETTI *et al.*, 2023).

Nesse contexto, é um desafio analisar dados não estruturados em documentos de patentes, especialmente em português. A maioria das técnicas de Reconhecimento de Entidades Nomeadas (REN) opera predominantemente em inglês e, embora possam ser adaptadas para o português, ainda apresentam grande margem para melhorias ao considerar as especificidades da língua (FERREIRA, 2019), especialmente quando integradas ao contexto educacional (SOUSA; MELLO, 2022).

O objetivo deste artigo é realizar uma análise computacional sobre a integração do conteúdo de patentes com o conteúdo dos planos de ensino. Nosso propósito é examinar as oportunidades, vantagens e desafios dessa integração, fornecendo recomendações práticas sobre os conhecimentos químicos que os alunos precisam aprimorar para abranger os requisitos químicos das patentes.

A principal contribuição deste artigo é a apresentação de uma metodologia inovadora para avaliar se um curso superior em química proporciona aos alunos os conhecimentos necessários para criar uma patente na área química, além de identificar em qual período do curso esses conhecimentos são adquiridos. Adicionalmente, a metodologia também identifica quais conteúdos precisam ser aprimorados pelos alunos para que possam desenvolver ou melhorar uma patente. Além disso, a disponibilidade de documentos com anotações químicas servirá de apoio para pesquisas futuras.

Este artigo segue a seguinte estrutura. Na Seção 2, revisamos alguns trabalhos relacionados. A Seção 3 apresenta a metodologia empregada para desenvolver a abordagem proposta. Em seguida, na Seção 4, detalhamos os experimentos realizados, os resultados obtidos e as limitações do trabalho. Finalmente, as conclusões e as direções futuras são abordadas na Seção 5.

2. Trabalhos Relacionados

Nesta seção, são apresentados estudos para a extração de informações de patentes, com o objetivo de identificar e classificar elementos específicos nos textos patentários, visando converter dados textuais não estruturados em informações estruturadas.

O trabalho de Akhondi *et al.* (2019) projetou um sistema automatizado que extrai e classifica a relevância de entidades químicas em patentes. Com um conjunto de dados contendo 18.789 anotações, 10% eram compostos relevantes, 88% irrelevantes e 2% ambíguos. Utilizando ferramentas proprietárias, o sistema alcançou uma medida F de 84% no reconhecimento de compostos no conjunto de desenvolvimento e 86% no de teste, enquanto a classificação de relevância obteve uma pontuação F de 86% no conjunto de desenvolvimento e 82% no de teste.

O trabalho de Izo e Oliveira (2023) propôs um sistema de Processamento de Linguagem Natural (PLN) para recuperar documentos relevantes aos planos de ensino, oferecendo suporte adicional aos estudos. Em seus experimentos, analisaram 15 planos de ensino e recuperaram materiais complementares de artigos científicos e patentes para



46,43% dos conteúdos identificados.

No trabalho de Puccetti *et al.* (2023) foi utilizado PLN para reconhecer as tecnologias mencionadas em uma patente. Os autores utilizaram o REN, comparando três métodos REN diferentes, medindo seus desempenhos em termos de *Precisão*, *Recall* e *Tempo computacional*. Foram testadas 1.600 patentes de quatro classes diversas de IPC e coletadas mais de 4.500 informações de tecnologias.

O estudo de Lima (2024) investigou o alinhamento entre a produção acadêmica das universidades e a proposição de patentes, utilizando dados do *Scival* e do *Instituto Nacional da Propriedade Industrial (INPI)*. Os resultados mostram que, apesar da significativa produção científica, poucas pesquisas se tornam patentes. Isso ocorre devido a fatores como a natureza das pesquisas, recursos e infraestrutura disponíveis, áreas de pesquisa predominantes e cultura institucional.

Todos os trabalhos citados são relevantes científica e academicamente, com metodologias que merecem consideração. No entanto, há uma lacuna na literatura em pesquisas que envolvam extração de informações químicas de documentos em português e na integração do conteúdo de patentes com o ensino em sala de aula. Dois estudos (IZO; OLIVEIRA, 2023; LIMA, 2024) exploraram a extração de informações químicas de patentes em português, com foco na integração educacional. Entretanto, (LIMA, 2024) analisou estatisticamente as produções acadêmicas, não examinando o conteúdo.

3. Metodologia

Para avaliar a abordagem proposta, foram estabelecidas várias etapas: inicialmente, o processo de coleta de dados será abordado, seguido pela análise dos dados coletados e pelas tecnologias empregadas.

3.1. Coleta de Dados:

Os dados são fundamentais para a validação de uma pesquisa. O procedimento de coleta de dados teve dois focos: o primeiro foi selecionar planos de ensino de um curso superior em química. O segundo foi selecionar patentes químicas em português.

A coleta de dados dos planos de ensino de componentes curriculares focou em um curso superior de química. Os dados foram selecionados do curso de graduação em Engenharia Química de uma instituição pública. Durante a análise, notou-se que nem todos os planos de ensino abordavam conteúdos diretamente relacionados à química. A seleção dos planos utilizáveis foi realizada em conjunto com um especialista na área. Planos de ensino de disciplinas como *Cálculo I*, *Fundamentos de Estatística* e *Introdução à Robótica* foram descartados por não estarem diretamente relacionados ao conteúdo químico, embora sejam importantes para a formação do discente.

Dos 64 planos de ensino, totalizando 3.375 horas, selecionamos 28 documentos com uma carga horária de 1.410 horas. Foram descartados 36 planos de ensino, totalizando 1.965 horas. A média foi de aproximadamente 3 planos de ensino por período, abrangendo todos os períodos, exceto o 10º. Os dados estatísticos sobre a seleção dos planos de ensino dos componentes curriculares estão apresentados na Tabela 1.

A coleta de dados de patentes químicas foi realizada por meio do acesso ao site do INPI. O critério inicial para a seleção das patentes foi que a Classificação Internacional de Patentes (IPC – *International Patent Classification*) estivesse relacionada à área química. O IPC é utilizado em mais de 100 países para auxiliar na busca e recuperação de documentos de patente, organizando-os de maneira a facilitar o acesso às informações tecnológicas e legais neles contidas (INPI, 2021). 30 documentos foram selecionados.

Com os 28 documentos de planos de ensino e os 30 documentos de patentes, o



Tabela 1. Coleta de Dados do Curso de Engenharia Química

Período	Comp. Curriculares Selecionados		Comp. Curriculares Descartados	
	Quantidade	CH	Quantidade	CH
1	3	120	4	255
2	3	150	4	210
3	5	240	2	120
4	2	150	5	240
5	3	180	4	210
6	3	180	4	210
7	3	150	4	240
8	4	150	4	240
9	2	90	4	210
10	0	0	1	30
Total	28	1.410	36	1.965

passo seguinte foi aplicar o algoritmo e validar a proposta de pesquisa.

3.2. Análise dos dados:

Após a coleta dos dados, foram realizadas várias etapas de pré-processamento. Inicialmente, os textos em formato *Portable Document Format* (PDF) foram convertidos para *Text Document* (TXT). Em seguida, cada sentença foi separada, colocando-as em linhas distintas.

Em trabalho anterior (IZO; OLIVEIRA, 2023) foi desenvolvido uma Gramática Local (GL – em inglês, *Local Grammar* (LG)) chamada LGq Plus, com o objetivo de anotar entidades químicas em português. A LGq Plus reconhece as seguintes categorias de entidades químicas: *Classes Químicas*, *Compostos Químicos*, *Elementos Químicos*, *Equipamentos Químicos* e *Métodos Químicos*. Desta forma, neste trabalho foi possível aprimorá-la, acrescentando novos recursos para a identificação de entidades químicas. A atualização passou a anotar:

- 15 novos *métodos químicos* e 18 novos *equipamentos químicos*;
- nova estrutura de *compostos químicos*, pois estes passaram a anotar além de seu próprio termo, a identificação da própria classificação funcional orgânica. Exemplos: ao anotar o composto químico *metanol*, o algoritmo também identifica e classifica esse composto como *Álcool*; Ao anotar o composto químico *1-Buteno*, o algoritmo também identifica e classifica esse composto como *Hidrocarboneto*.
- duas novas categorias de entidades químicas: *Conceitos Químicos* e *Reações Químicas*.

Assim, a LGq Plus passou a reconhecer 7 categorias: *Classes Químicas*, *Compostos Químicos*, *Conceitos Químicos*, *Elementos Químicos*, *Equipamentos Químicos*, *Métodos Químicos* e *Reações Químicas*.

Ao final desta etapa, obtivemos 28 planos de ensino e 30 patentes, todos anotados com as entidades químicas de todas as categorias. A Tabela 2 apresenta o total de entidades identificadas e anotadas nos documentos de patentes e planos de ensino.

3.3. Aplicação prática

Para nossa abordagem foram utilizadas algumas aplicações:

- *Extração do texto*: Utilizamos o Apache Tika¹ para extrair o texto dos arquivos em formato PDF.

¹<https://tika.apache.org/>



Tabela 2. Entidades Químicas Identificadas e Anotadas em Documentos de Patentes e Planos de Ensino

Categoria	Patentes	Planos de Ensino
Classes Químicas	379	10
Compostos Químicos	618	3
Conceitos Químicos	2.252	216
Elementos Químicos	1.438	9
Equipamentos Químicos	743	8
Métodos Químicos	940	73
Reações Químicas	731	42
Total	7.101	361

- Aplicação da LG: A LG se baseia em grafos autômatos e após definí-los, utilizamos a linguagem *Shell Script* para chamar as ferramentas *Unitex/GramLab*, aplicando os grafos e anotando as entidades nos documentos utilizados.
- Processamento de análise dos dados: A análise estatística das entidades identificadas e anotadas nas patentes e a integração dos dados com o plano de ensino constituem um processo complexo. Para isso, utilizamos um algoritmo desenvolvido em *Python*, aproveitando bibliotecas específicas da linguagem para processamento matemático.
- Interface Usuário Máquina: A interface que permite a interação do usuário com o algoritmo foi desenvolvida através de plataforma *web*. Esta funcionalidade foi implementada utilizando tecnologia *HTML5* juntamente com a linguagem *PHP*.

Com os dados coletados, identificados e anotados, iniciou-se a etapa dos experimentos, que será detalhada na Seção 4.

4. Resultados e Experimentos

Os experimentos são fundamentais para a pesquisa científica, permitindo a análise e validação dos resultados. Como parte da pesquisa aplicada, elaboramos uma série de experimentos para verificar a viabilidade de integrar o conteúdo dos componentes curriculares com o conhecimento necessário para a elaboração e melhoria de patentes. Esta seção descreve a condução dos experimentos, detalhando o planejamento, os resultados e as discussões, e, por fim, relatando as limitações deste projeto.

4.1. Planejamento dos experimentos

Conforme já mencionado na Seção 3, foram selecionados 28 planos de ensino de um curso de graduação em Engenharia Química de uma instituição pública de ensino, além de 30 patentes químicas do INPI. Esse conjunto de dados foi utilizado nos experimentos.

O objetivo dos experimentos foi analisar computacionalmente os dados para:

1. Determinar se um curso superior em química proporciona aos alunos os conhecimentos necessários para criar uma patente e identificar em qual período do curso esses conhecimentos são adquiridos.
2. Identificar quais conteúdos precisam ser aprimorados pelos alunos para que eles possam desenvolver ou melhorar uma patente.

Para atingir tais objetivos, foram definidas duas fases dos experimentos.

Número da Patente: **patent01-BR-102017021115-0-B1**Título da Patente: **FORMULAÇÃO DESINFETANTE PARA DESINFECÇÃO A SECO DE AMBIENTES**

Total de Entidades Diferentes: 15 | Total de Entidades: 43

#	Período	Componente Curricular	Métrica Base		Métrica com Peso (Frequência)	
			%R %A	% Acumulada	%R %A	% Acumulada
PE2	1	Química Geral I	46,69 46,69	46,69	74,43 74,43	74,43
PE3	1	Química Geral Experimental	13,34 6,67	53,36	6,98 4,65	79,08
PE4	2	Balanço de Massa e Energia	0,00 0,00	53,36	0,00 0,00	79,08
PE5	2	Química Geral II	13,34 0,00	53,36	32,56 0,00	79,08
PE6	2	Química Orgânica I	6,67 0,00	53,36	27,91 0,00	79,08
PE7	3	Fundamentos de Bioquímica	6,67 0,00	53,36	27,91 0,00	79,08
PE8	3	Química Orgânica II	13,34 6,67	60,03	30,24 2,33	81,41
PE9	3	Química Orgânica Experimental	6,67 6,67	66,70	6,98 6,98	88,39
PE10	3	Físico-Química I	6,67 6,67	73,37	2,33 2,33	90,72
PE11	3	Química Inorgânica	6,67 0,00	73,37	27,91 0,00	90,72
PE12	4	Físico-Química II	20,01 0,00	73,37	13,96 0,00	90,72
PE13	4	Química Analítica	26,68 0,00	73,37	44,19 0,00	90,72
PE14	5	Análise Instrumental	0,00 0,00	73,37	0,00 0,00	90,72
PE15	5	Físico-Química Experimental	6,67 0,00	73,37	4,65 0,00	90,72
PE16	5	Termodinâmica	0,00 0,00	73,37	0,00 0,00	90,72
PE17	6	Operações Unitárias I	6,67 0,00	73,37	4,65 0,00	90,72
PE18	6	Laboratório de Engenharia Química I	13,34 6,67	80,04	6,98 2,33	93,05
PE19	6	Ciência e Tecnologia de Materiais	6,67 0,00	80,04	2,33 0,00	93,05
PE20	7	Operações Unitárias II	6,67 0,00	80,04	4,65 0,00	93,05
PE21	7	Laboratório de Engenharia Química II	6,67 0,00	80,04	4,65 0,00	93,05
PE24	8	Laboratório de Engenharia Química III	6,67 0,00	80,04	4,65 0,00	93,05
PE25	8	Operações Unitárias III	6,67 0,00	80,04	6,98 0,00	93,05
PE26	8	Processos Industriais I	6,67 0,00	80,04	20,93 0,00	93,05

* %R: porcentagem Real de termos identificados.

** %A: porcentagem Aproveitada de termos que ainda não haviam sido adicionados em períodos anteriores.

*** Em vermelho os componentes curriculares sem termos químicos identificados para esta patente.

Figura 1. Análise Computacional dos Dados Químicos Referentes a Patente BR-102017021115-0-B1.

ELEMENTOS QUÍMICOS: • Cloro (9) • Sódio (4) • Cálcio (3) • Potássio (1)	CLASSES QUÍMICAS: • Aminas (1) • Imidas (1) • Iminas (1)	MÉTODOS QUÍMICOS: • Fusão (1) • Suspensão (1)
CONCEITOS QUÍMICOS: • ácido (12) • Solvente (3) • Solução (2) • Óxido (2) • Densidade (1)	REAÇÕES QUÍMICAS: • Oxidação (1)	

* em parênteses está o número de vezes que o termo apareceu.

** Em vermelho os termos não identificados em nenhum plano de ensino.

Figura 2. Relatório dos Conhecimentos Químicos Identificados na Patente BR-102017021115-0-B1.

4.2. Experimentos Fase 1 (EF1)

Nesta fase, cada patente foi processada individualmente, com as entidades sendo identificadas e anotadas. Em seguida, os dados foram integrados aos planos de ensino. Por fim, foi desenvolvido um gráfico estatístico, classificando os termos químicos por ocorrência, para apresentar os resultados aos usuários. Este experimento gerou as seguintes informações:

- Planos de Ensino sem nenhuma entidade química identificada e anotada.
- Análise individual de cada plano de ensino, indicando o percentual de conteúdo abordado relacionado à patente analisada (Ver Figura 1).
- Percentual de Métrica Base (MB) e Métrica com Peso (MP) dos conteúdos cobertos pelo plano de ensino relacionados à patente analisada (Ver Figura 1).
- Quantidade de termos químicos encontrados e anotados na patente analisada (Ver Figura 2).



- Indicação de quais conteúdos precisam ser aprimorados pelos alunos para que eles possam desenvolver ou melhorar uma patente (Ver Figura 2).

Na Figura 1 é apresentado um exemplo dos dados que são gerados para o usuário referente a patente *BR-102017021115-0-B1*. A Figura 2 mostra um exemplo dos conteúdos que precisam ser aprimorados pelos alunos para desenvolver ou melhorar essa mesma patente.

É importante conceituar alguns parâmetros utilizados nesse experimento. As métricas propostas foram *Métricas Base (MB)* e *Métricas por Peso (MP)*. A MB foi calculada tendo como base uma divisão simples de *1 ocorrência* dividido pelo *total de termos químicos diferentes (types)* encontrados no documento de patente. Logo, cada termo diferente encontrado no plano de ensino teria o mesmo valor MB, de acordo com (1). Em um exemplo prático, utilizando como base a Figura 2, o MB de cada um dos 15 termos químicos diferentes será: $1 / 15 \approx 6,67\%$.

A MP foi calculada tendo como base o cálculo da *Term Frequency (TF)*. Há várias noções diferentes associadas à frequência, mas, quando usada sem especificações adicionais, ela geralmente se refere ao número de ocorrências de uma expressão em um documento (FERREIRA; LOPES, 2019). Nesta pesquisa o objetivo era descobrir a frequência absoluta. Assim, houve a divisão do *número de ocorrências de um determinado termo t* pelo *total de termos do documento d* (tokens). Logo, cada termo diferente encontrado no plano de ensino terá um valor diferente para MP, de acordo com (2). Em um exemplo prático, utilizando como base a Figura 2, o MP do Elemento Químico *cloro* será $9 / 43 \approx 20,93\%$. E assim por diante. O MP do Método Químico *fusão* será $1 / 43 \approx 2,32\%$. E assim por diante. Os valores de MB e MP são multiplicados por 100 para facilitar a leitura e analisarmos porcentagens entre 0 e 100%.

$$MB_{t,d} = \frac{1}{types} \quad (1) \quad MP_{t,d} = \frac{frequência\ absoluta\ t}{tokens} \quad (2)$$

Em cada uma das métricas MB e MP foi calculado a Porcentagem Real (%R) e a Porcentagem Aproveitada (%A). A %R é o equivalente a calcular a porcentagem de termos real localizadas em determinado documento. A %A extrai de %R somente os termos que ainda não foram identificados em documentos anteriores. Na Figura 1, por exemplo, a %R e a %A das métricas do PE2 são iguais pois não houve a análise de nenhum plano de ensino anteriormente. Na análise do PE3 a %R da MB é de 13,34, porém foi aproveitado (%A) 6,67 dos termos, pois o restante já havia sido identificado no plano PE2.

4.3. Experimentos Fase 2 (EF2)

Nesta segunda fase de experimentos, o objetivo foi realizar testes com todas as 30 patentes para identificar quais são suficientemente cobertas pelos conteúdos dos planos de ensino. Essa análise é importante para obter uma visão geral das patentes e extrair dados através de análise computacional sobre quais tipos de entidades são mais frequentemente identificadas e anotadas em patentes que têm uma cobertura significativa dos conteúdos de um curso de química.

4.4. Resultados e Limitações

Durante EF1, foi observado que os planos de ensino de *Introdução à Engenharia Química* (1º Período), *Tratamento de Águas* (7º Período), *Tratamento de Efluentes* (8º Período), *Controle de Processos* (9º Período) e *Processos Industriais II* (9º Período) não encontraram conteúdo químico em nenhuma das 30 patentes, representando 17,85% dos



documentos. No entanto, esses planos não foram descartados, pois poderão ser úteis em futuras melhorias na LG_q Plus.

Em relação aos planos de ensino que identificaram conteúdos químicos, o PE4 e o PE10 observaram apenas um termo químico cada. O PE13 identificou 81 termos químicos. A média de termos químicos por plano de ensino foi de 15,69, e a mediana foi de 9.

Após a EF1, preparamos a integração e análise simultânea de todas as 30 patentes. Assim, geramos a EF2, que está parcialmente (Patente 1 a Patente 11) demonstrada na Figura 3. Na primeira linha tem-se as patentes de 1 a 30 (parcialmente cortada nesta figura). Na segunda linha os rótulos das informações analisadas computacionalmente. A primeira coluna se refere ao período (Per) em que o plano de ensino (PE) da segunda coluna é ministrado no curso. As demais colunas seguintes são referentes às métricas MB e MP de cada uma das patentes. Nesta figura destacamos em vermelho em que período/plano de ensino foi observado acima de 70% do conhecimento químico referente a patente. Após analisar os planos de ensino do todo o curso de Engenharia Química, a MP se sobressaiu em relação a 17 planos de ensino, contra 13 da MB.

		Pat1		Pat2		Pat3		Pat4		Pat5		Pat6		Pat7		Pat8		Pat9		Pat10		Pat11	
Per	PE	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP	MB	MP
1	2	46,69	74,43	38,45	40,00	32,64	12,37	50,04	32,70	50,04	59,49	15,21	14,13	19,17	23,10	31,56	6,27	26,10	11,89	24,66	54,85	50,00	73,52
1	3	53,36	79,08	46,14	41,29	34,68	19,55	58,38	44,56	54,21	64,40	18,59	14,88	21,30	23,43	36,82	9,85	30,45	14,12	24,66	54,85	68,75	82,34
2	4	53,36	79,08	46,14	41,29	34,68	19,55	58,38	44,56	54,21	64,40	18,59	14,88	21,30	23,43	36,82	9,85	30,45	14,12	24,66	54,85	68,75	82,34
2	5	53,36	79,08	61,52	42,59	36,72	20,12	58,38	44,56	58,38	65,63	21,97	15,82	27,69	26,40	42,08	12,24	39,15	18,21	28,77	56,51	68,75	82,34
2	6	53,36	79,08	61,52	42,59	38,76	20,41	58,38	44,56	58,38	65,63	21,97	15,82	27,69	26,40	42,08	12,24	39,15	18,21	28,77	56,51	68,75	82,34
3	7	53,36	79,08	61,52	42,59	38,76	20,41	58,38	44,56	58,38	65,63	21,97	15,82	27,69	26,40	42,08	12,24	39,15	18,21	28,77	56,51	68,75	82,34
3	8	60,03	81,41	61,52	42,59	38,76	20,41	58,38	44,56	62,55	66,24	23,66	16,01	29,82	27,06	42,08	12,24	43,50	18,58	30,14	56,79	68,75	82,34
3	9	66,70	88,39	61,52	42,59	44,88	33,06	58,38	44,56	66,72	71,15	25,35	16,39	36,21	36,96	47,34	12,54	47,85	19,70	32,88	57,35	75,00	85,28
3	10	73,37	90,72	61,52	42,59	44,88	33,06	58,38	44,56	66,72	71,15	27,04	17,14	36,21	36,96	47,34	12,54	47,85	19,70	32,88	57,35	75,00	85,28
3	11	73,37	90,72	61,52	42,59	44,88	33,06	58,38	44,56	66,72	71,15	27,04	17,14	36,21	36,96	47,34	12,54	47,85	19,70	32,88	57,35	75,00	85,28
4	12	73,37	90,72	61,52	42,59	44,88	33,06	58,38	44,56	66,72	71,15	27,04	17,14	36,21	36,96	47,34	12,54	47,85	19,70	32,88	57,35	75,00	85,28
4	13	73,37	90,72	61,52	42,59	44,88	33,06	63,94	46,25	66,72	71,15	27,04	17,14	38,34	40,92	57,86	21,80	47,85	19,70	35,62	58,73	75,00	85,28
5	14	73,37	90,72	61,52	42,59	44,88	33,06	63,94	46,25	66,72	71,15	27,04	17,14	38,34	40,92	57,86	21,80	47,85	19,70	35,62	58,73	75,00	85,28
5	15	73,37	90,72	61,52	42,59	44,88	33,06	63,94	46,25	66,72	71,15	27,04	17,14	38,34	40,92	57,86	21,80	47,85	19,70	35,62	58,73	75,00	85,28
5	16	73,37	90,72	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	27,04	17,14	38,34	40,92	63,12	50,76	56,55	39,77	35,62	58,73	75,00	85,28
6	17	73,37	90,72	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	28,73	17,33	40,47	41,91	63,12	50,76	60,90	40,14	36,99	59,56	81,25	91,16
6	18	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	28,73	17,33	40,47	41,91	63,12	50,76	60,90	40,14	36,99	59,56	81,25	91,16
6	19	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	63,12	50,76	60,90	40,14	36,99	59,56	81,25	91,16
7	20	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	68,38	51,06	60,90	40,14	36,99	59,56	81,25	91,16
7	21	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	68,38	51,06	60,90	40,14	36,99	59,56	81,25	91,16
8	24	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	68,38	51,06	60,90	40,14	36,99	59,56	81,25	91,16
8	25	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	68,38	51,06	60,90	40,14	36,99	59,56	81,25	91,16
8	26	80,04	93,05	69,21	69,04	44,88	33,06	66,72	52,46	66,72	71,15	30,42	17,71	40,47	41,91	68,38	51,06	60,90	40,14	36,99	59,56	81,25	91,16

Figura 3. Parte da Análise Computacional dos Dados Químicos das 30 Patentes e dos 28 Planos de Ensino

A menor taxa de porcentagem observada foi na patente *BR-112019004910-0-B1* (PAT6), apresentando 30,42% de MB e 17,71% de MP. Ao analisar minuciosamente esta patente, percebe-se que a mesma possui 531 termos químicos observados, dos quais 59 termos eram distintos. Detalhando um pouco mais, observou-se que dos 59 termos distintos, 27 são relacionados a compostos químicos, totalizando 163 *tokens*. Este resultado foi importante, pois apesar de não identificar tais compostos nos planos de ensino, o algoritmo indicou que o aluno deveria se aperfeiçoar em tais conhecimentos para conseguir melhorar ou analisar a PAT6.

A média de conhecimento adquirido entre as 30 patentes foi atingida ao final do terceiro período, com MB marcando 52% e MP marcando 50% dos conteúdos. Isso demonstra que é possível utilizar o conhecimento adquirido durante o curso para analisar e criar e melhorar patentes químicas. Além dos conhecimentos explícitos identificados



nesta análise computacional, o algoritmo aponta quais conhecimentos químicos precisam ser aprimorados pelos alunos para atingir 100% dos conhecimentos químicos requeridos pela patente, conforme ilustrado na Figura 2.

Dividiu-se os resultados em faixas percentuais, pois esta técnica pode aumentar a precisão e interpretabilidade das análises estatísticas na pesquisa educacional (YORK; GIBSON; RANKIN, 2019). As faixas percentuais de MB e MP estão representadas na Tabela 3. A maioria das patentes (14) apresentou valores de MB na faixa de 50-70%. Para MP, foram observadas 9 patentes tanto na faixa de 30-50% quanto na faixa de 50-70%. Três patentes obtiveram valores de MP acima de 90%. Analisando essas patentes, observa-se que elas têm em comum uma quantidade significativa de termos químicos nas categorias: *Conceitos Químicos*, *Elementos Químicos*, *Métodos Químicos* e *Reações Químicas*. Ao analisar outras patentes, observa-se que os conteúdos químicos mais bem alinhados são, de fato, aqueles das categorias mencionadas, como também é evidenciado na Tabela 2.

Tabela 3. Faixa Percentual das Métrica MB e MP em Relação as 30 Patentes.

	10-30	30-50	50-70	70-90	90-100
MB	0	10	14	6	0
MP	2	9	9	7	3

Embora a análise computacional dos dados de patentes e planos de ensino tenha demonstrado a possibilidade de integrar essas informações para construir conhecimentos tanto para discentes quanto para docentes, algumas limitações deste trabalho foram identificadas. Os planos de ensino geralmente são estruturados com seções como *Ementa*, *Objetivos*, *Conteúdos*, entre outras, mas o texto inserido neles não segue um padrão bem definido. A maioria dos textos dos planos de ensino tende a ser mais generalista, enquanto as patentes contêm textos mais específicos.

Consequentemente, alguns conteúdos abordados em sala de aula podem não ser identificados pelo algoritmo durante o processamento. Além disso, os professores podem discutir conteúdos extras ou tópicos de estágio com os alunos, que não são detalhados nos planos de ensino. Todo esse conhecimento adicional seria relevante para o algoritmo, mas acaba não sendo capturado na análise computacional quando analisamos somente os planos de ensino disponibilizados pelo professor.

5. Conclusões e Trabalhos Futuros

As patentes são fontes valiosas de informação científica, destacando inovações e conhecimentos técnicos obtidos através de pesquisas. Utilizar patentes como fonte de aprendizagem é uma estratégia enriquecedora com um enorme potencial educacional. No entanto, é perceptível que poucas estruturas curriculares exploram esses documentos científicos de maneira adequada.

Neste trabalho, apresentamos uma metodologia para integrar os planos de ensino de um curso superior em Engenharia Química com documentos de patentes. Nosso objetivo é fornecer recomendações práticas sobre os conhecimentos químicos que os alunos precisam aprimorar durante o curso, a fim de atender aos requisitos técnicos das patentes. Além disso, a metodologia proposta permite recuperar informações que facilitam o acompanhamento gradual da evolução da integração entre os conteúdos abordados em cada período do curso e os conhecimentos presentes nas patentes. Esse tipo de trabalho também possibilita que professores ou membros do colegiado acrescentem atividades extras ou adaptem seus planos de ensino, preparando melhor os alunos para



criar ou melhorar patentes.

Os resultados demonstraram que é viável integrar patentes à estrutura curricular de cursos superiores. Em trabalhos futuros, pretende-se investigar outros tipos de documentos presentes na estrutura dos cursos superiores, tais como planos de aula, relatórios de atividades complementares e projetos extracurriculares disponíveis aos alunos, e desta forma sanar algumas limitações deste trabalho. Também pretende-se analisar quais equipamentos são utilizados em disciplinas como Laboratório de Engenharia Química I (PE18), II (PE21) e III (PE24). Ter essa relação pode facilitar e aumentar o número de *Equipamentos Químicos* observados nos planos de ensino.

Referências

- AKHONDI, S. A. *et al.* Automatic identification of relevant chemical compounds from patents. **Database**, v. 2019, p. baz001, 01 2019. ISSN 1758-0463.
- FERREIRA, J. D. C. **Python para Pré-processamento e Extração de Características a partir de Texto Português**. Dissertação (Mestrado) — Universidade de Coimbra, 2019.
- FERREIRA, M.; LOPES, M. **Para Conhecer: linguística computacional**. São Paulo: Editora Contexto, 2019. 192 p.
- GREENBERG, A.; COHEN, A.; GREWAL, M. Patent landscape of brain-machine interface technology. **Nature Biotechnology**, Nature Publishing Group US New York, v. 39, n. 10, p. 1194–1199, 2021.
- HE, J. *et al.* Chemu 2020: Natural language processing methods are effective for information extraction from chemical patents. **Frontiers in Research Metrics and Analytics**, v. 6, 2021. ISSN 2504-0537.
- INPI. **Instituto Nacional de Propriedade Industrial: Manual básico para proteção por patentes de invenções, modelos de utilidade e certificados de adição**. 2021.
- IZO, F.; OLIVEIRA, E. Recuperação automatizada de documentos científicos por meio da análise de planos de ensino. **Revista Novas Tecnologias na Educação**, v. 21, n. 2, p. 168–177, 2023.
- KALTHAUS, M. Knowledge recombination along the technology life cycle. **Journal of Evolutionary Economics**, Springer, v. 30, n. 3, p. 643–704, 2020.
- KRESTEL, R.; CHIKKAMATH, R.; HEWEL, C.; RISCH, J. A survey on deep learning for patent analysis. **World Patent Information**, v. 65, p. 102035, 2021. ISSN 0172-2190.
- LIMA, C. d. C. **Interação entre a produção científica e os dados de patentes das universidades federais brasileiras**. 98 p. Dissertação (Mestrado Profissional em Economia) — Universidade de Brasília, Brasília, 2024.
- PUCETTI, G.; GIORDANO, V.; SPADA, I.; CHIARELLO, F.; FANTONI, G. Technology identification from patent texts: A novel named entity recognition method. **Technological Forecasting and Social Change**, v. 186, p. 122160, 2023. ISSN 0040-1625.
- SAAD, F. Named entity recognition for biomedical patent text using bi-lstm variants. In: **Proceedings of the 21st International Conference on Information Integration and Web-Based Applications & Services**. New York, NY, USA: Association for Computing Machinery, 2020. (iiWAS2019), p. 617–621. ISBN 9781450371797.
- SOUSA, E. B. G. de; MELLO, R. F. Aplicação de reconhecimento de entidades nomeadas para análise automática de textos narrativos em produções textuais do ensino fundamental. In: SBC. **Anais Estendidos do XI Congresso Brasileiro de Informática na Educação**. Manaus, 2022. p. 150–155.
- YORK, T. T.; GIBSON, C.; RANKIN, S. Defining and measuring academic success. **Practical assessment, research, and evaluation**, v. 20, n. 1, p. 5, 2019.