

Inteligência Artificial Generativa Aplicada na Análise da Produção Científica

Programa de Pós-Graduação em Informática e Gestão do Conhecimento (PPGI) - Uninove
Programa de Pós-Graduação em Gestão de Projetos (PPGP) - Uninove

Dr. Edson Melo de Souza

AULA 03

Fontes e APIs de Dados Científicos

Objetivos

- ▶ Analisar comparativamente as bases de dados (Teórico-Crítico).
- ▶ Interpretar a estrutura de metadados científicos (Conceitual).
- ▶ Operacionalizar a extração de dados via API (Prático-Técnico).

Fundamentos e os “Gigantes” Proprietários

Fundamentos e os “Gigantes” Proprietários

Contextualização

- ▶ O Ecossistema da Informação Científica:
 - ▶ **Editoras:** Elsevier, Springer, Wiley, PubMed ([CULBERT *et al.*, 2024](#)).
 - ▶ **Indexadores:** Web of Science, Scopus, Google Scholar ([VISSER *et al.*, 2021](#)).
 - ▶ **Produção Científica:** Artigos, livros, dados, códigos.
 - ▶ **Repositórios e Bases de Dados:** Armazenamento e acesso (OpenAlex).
- ▶ Para que servem as bases de dados científicas? ([KAICINOVA; PAVELEK, 2025](#))
 - ▶ Pesquisa e revisão bibliográfica.
 - ▶ Análise de citações e métricas.
 - ▶ Extração de dados para estudos bibliométricos.

Conceitos Chave

Bases de Dados Científicas

- ▶ Meadados vs. Texto Completo.
- ▶ Identificadores Únicos: DOI, ORCID, ISSN, ISBN.
 - ▶ **DOI** (Digital Object Identifier): Identificador persistente para documentos digitais.
 - ▶ **ORCID** (Open Researcher and Contributor ID): Identificador único para autores.
 - ▶ **ISSN** (International Standard Serial Number): Identificador para publicações seriadas.
 - ▶ **ISBN** (International Standard Book Number): Identificador para livros.
- ▶ APIs (Application Programming Interfaces):
 - ▶ RESTful APIs.
 - ▶ Formatos de Dados: JSON, XML, CSV.

Principais Bases de Dados Científicas

Comparativo

- ▶ **Scopus (Elsevier)**: Cobertura multidisciplinar, métricas de impacto, API paga.
- ▶ **Web of Science (Clarivate)**: Cobertura multidisciplinar, métricas de impacto, API paga.
- ▶ **Google Scholar**: Acesso gratuito, cobertura ampla, sem API oficial.
- ▶ **PubMed**: Foco em ciências da saúde, acesso gratuito, API robusta.
- ▶ **CrossRef**: Metadados de publicações acadêmicas, API aberta.
- ▶ **Dimensions**: Ampla cobertura multidisciplinar, API paga com opções gratuitas.
- ▶ **Semantic Scholar**: IA para análise de literatura, API gratuita.
- ▶ **OpenAlex**: Base de dados aberta, API gratuita e acessível ([PRIEM et al., 2022](#)).

- ▶ **Cobertura Multidisciplinar:** Ciências, Tecnologia, Medicina, Ciências Sociais, Artes e Humanidades.
- ▶ **Metadados Abrangentes:** Títulos, autores, afiliações, resumos, citações.
- ▶ **Ferramentas de Análise:** Métricas de citação, análise de coautoria, tendências de pesquisa.
- ▶ **Métricas de Impacto:** Fator de impacto, índice h, citações por artigo.
- ▶ **Métricas Associadas:** Altmetrics, PlumX, CiteScore, SJR, SNIP.
- ▶ **Acesso via API:** Requer assinatura institucional, documentação detalhada.

- ▶ **Cobertura Multidisciplinar:** Ciências, Tecnologia, Medicina, Ciências Sociais, Artes e Humanidades.
- ▶ **Metadados Abrangentes:** Títulos, autores, afiliações, resumos, citações.
- ▶ **Ferramentas de Análise:** Métricas de citação, análise de coautoria, tendências de pesquisa.
- ▶ **Métricas de Impacto:** Fator de impacto, índice h, citações por artigo.
- ▶ **Métricas Associadas:** Altmetrics, PlumX, Journal Citation Reports.
- ▶ **Acesso via API:** Requer assinatura institucional, documentação detalhada.

A Revolução Aberta e Estrutura de Dados

OpenAlex

OpenAlex - O Novo Player

Visão Geral

O OpenAlex é uma grande base de dados acadêmica aberta, gratuita e global, que reúne informações sobre artigos científicos, autores, instituições, periódicos e áreas de pesquisa ([ARIA et al., 2024](#)).

- ▶ **Cobertura Abrangente:** Milhões de artigos, autores e instituições.
- ▶ **Metadados Detalhados:** Títulos, resumos, citações, referências.
- ▶ **Acesso via API:** Gratuito e bem documentado.
- ▶ **Formato de Dados:** JSON, fácil de integrar com outras ferramentas.
- ▶ **Comunidade Ativa:** Suporte e desenvolvimento contínuo.

OpenAlex - Estrutura de Dados

Entidades Principais

- ▶ **Works (Trabalhos)**: Artigos, livros, capítulos.
- ▶ **Authors (Autores)**: Pesquisadores e colaboradores.
- ▶ **Institutions (Instituições)**: Universidades, centros de pesquisa.
- ▶ **Journals (Periódicos)**: Revistas científicas.
- ▶ **Concepts (Conceitos)**: Áreas de pesquisa e tópicos.

Comparativo Técnico

Característica	Scopus	Web of Science	OpenAlex
Cobertura	Multidisciplinar	Multidisciplinar	Multidisciplinar
Acesso	Pago	Pago	Gratuito
API	Sim (paga)	Sim (paga)	Sim (gratuita)
Formato de Dados	JSON, XML	JSON, XML	JSON
Metadados	Abrangentes	Abrangentes	Abrangentes
Métricas	Sim	Sim	Limitadas

Tabela: Comparativo Técnico entre Scopus, Web of Science e OpenAlex

Intervalo de 15 Minutos

- ▶ **O que é uma API RESTful?**

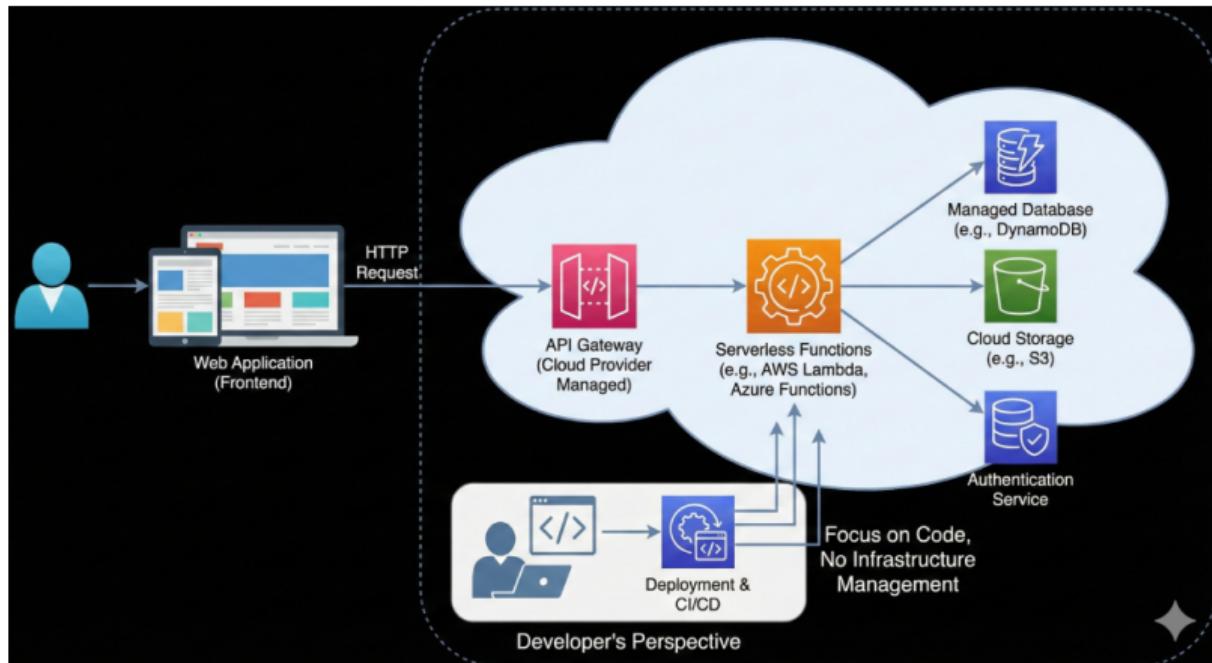
- ▶ Interface que permite a comunicação entre sistemas via HTTP.
- ▶ Utiliza métodos HTTP: GET, POST, PUT, DELETE.

- ▶ **Formato de Dados Comum:** JSON (JavaScript Object Notation).

- ▶ **Componentes Principais:**

- ▶ Endpoint: URL base para acessar a API.
- ▶ Recursos: Entidades disponíveis (ex.: works, authors).
- ▶ Parâmetros: Filtros e opções para refinar consultas.

Funcionamento de APIs RESTful



Extração de Dados via API do OpenAlex

- ▶ **Documentação da API:** <https://docs.openalex.org/>
- ▶ **Exemplo de Requisição:** Buscar artigos sobre “Inteligência Artificial”.
 - ▶ Endpoint: `https://api.openalex.org/works?filter=title.search:"Biologia"`
- ▶ **Parâmetros Comuns:**
 - ▶ `filter`: Filtrar resultados por critérios específicos.
 - ▶ `sort`: Ordenar resultados (ex.: por data, relevância).
 - ▶ `per-page`: Número de resultados por página.
 - ▶ `page`: Página específica dos resultados.
- ▶ **Exemplo de Resposta:** JSON com metadados dos artigos encontrados.

- ▶ **Leitura 1:** Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic
https://doi.org/10.1162/qss_a_00112
- ▶ **Leitura 2:** OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts
<https://arxiv.org/abs/2205.01833>
- ▶ **Leitura 3:** Reference coverage analysis of OpenAlex compared to Web of Science and Scopus
<https://arxiv.org/abs/2401.16359>

Resumo

Nesta aula foram apresentadas informações sobre fontes de dados científicos e APIs, comparando bases proprietárias (Scopus e Web of Science) com a plataforma aberta OpenAlex. O texto detalha o funcionamento técnico de APIs RESTful e orienta sobre a extração automatizada de metadados acadêmicos, utilizando os endpoints e filtros da API do OpenAlex para fins de pesquisa e bibliometria.

- ▶ Bases de dados científicas são essenciais para pesquisa e análise.
- ▶ Scopus e Web of Science são líderes proprietários, enquanto OpenAlex é uma alternativa aberta.
- ▶ APIs RESTful permitem extração automatizada de dados científicos.
- ▶ OpenAlex oferece acesso gratuito e fácil integração via API.

Bases de Dados Científicas: Gigantes Proprietários vs. A Revolução Aberta

Gigantes Proprietários: Scopus & Web of Science

Cobertura Amplia e Multidisciplinar
Abrangem ciências, tecnologia, medicina, ciências sociais, artes e humanidades.



Métricas de Impacto Robustas
Offerem análises detalhadas como fator de impacto, índice h e CiteScore.



Acesso Restrito e Pago
O acesso completo aos dados e à API requer uma assinatura institucional.



A Revolução Aberta: OpenAlex

Base de Dados Aberta e Gratuita
Retine milhões de artigos, autores e instituições de forma acessível e global.



API Gratuita e Acessível
Permite a extração de dados automatizada com uma API bem documentada.



Comparação Técnica

Característica	Scopus	Web of Science	OpenAlex
Acesso	Pago	Pago	Gratuito
API	Sim (paga)	Sim (paga)	Sim (gratuita)
Formato de Dados	JSON, XML	JSON, XML	JSON

Estrutura de Dados Organizada
Organiza a informação em 5 entidades:
Trabalhos, Autores, instituições, Periódicos e Conceitos.



ATIVIDADE AVALIATIVA 1

(10% da nota final)

As instruções para esta atividade estão disponíveis no Google Classroom na guia ATIVIDADES → ATIVIDADE 1.

Referências

- ARIA, M. et al. **openalexR: An R-Tool for Collecting Bibliometric Data from OpenAlex.** *The R Journal*, 2024. doi:[10.32614/RJ-2023-089](https://doi.org/10.32614/RJ-2023-089).
- CULBERT, J. et al. **Reference Coverage Analysis of OpenAlex compared to Web of Science and Scopus.** *arXiv preprint arXiv:2401.16359*, 2024. doi:[10.48550/arXiv.2401.16359](https://doi.org/10.48550/arXiv.2401.16359).
- KAICINOVA, V.; PAVELEK, L. **Generative Artificial Intelligence and the Media: A Bibliometric Analysis in Academic Research.** *Media Literacy and Academic Research*, v. 8, n. 1, p. 29–47, 2025. doi:[10.34135/mlar-25-01-02](https://doi.org/10.34135/mlar-25-01-02).
- PRIEM, J.; PIWOWAR, H.; ORR, R. **OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts.** *arXiv preprint arXiv:2205.01833*, 2022. doi:[10.48550/arXiv.2205.01833](https://doi.org/10.48550/arXiv.2205.01833).
- VISSEER, M.; ECK, N. J. van; WALTMAN, L. **Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic.** *Quantitative Science Studies*, MIT Press, v. 2, n. 1, p. 20–41, 2021.