

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas – Departamento de Ciência da Computação

Mapeamento do Ecossistema de Inovação em Minas Gerais

Tipo de Pesquisa: **Mista (Científica e Tecnológica)**

Orientador: Marcos André Gonçalves

Proposta de Trabalho de Conclusão de Curso – 2025

Autor: Renato Silva Santos

Belo Horizonte – MG
2024

1 Introdução

Este documento apresenta a proposta de um Projeto Orientado em Computação (POC) em Ciência da Computação a ser desenvolvido ao longo do ano de 2025 na Universidade Federal de Minas Gerais (UFMG). O tema central é o mapeamento do ecossistema de desenvolvimento e inovação no estado de Minas Gerais, com foco em tecnologias, áreas de pesquisa, startups e Instituições de Ciência e Tecnologia (ICTs) mineiras. Observa-se que as informações sobre esses atores estão dispersas em diferentes fontes e formatos, dificultando uma visão integrada do panorama de inovação regional.

O problema a ser investigado consiste em como traçar um perfil abrangente da inovação tecnológica em Minas Gerais conectando pesquisadores acadêmicos com iniciativas empreendedoras (startups e empresas de base tecnológica) e instituições de pesquisa. Atualmente, não existe uma base de dados unificada que agregue informações sobre quais tecnologias e áreas de pesquisa estão em destaque no estado, quais startups atuam nesses segmentos e quais pesquisadores contribuem para essas frentes de inovação. Essa falta de integração dificulta tanto a identificação de oportunidades de colaboração entre academia e mercado quanto o direcionamento de políticas de incentivo à inovação.

A justificativa para este projeto reside na importância de compreender e fortalecer o ecossistema de inovação de Minas Gerais. O estado abriga polos tecnológicos emergentes e um grande número de atores inovadores. Segundo um estudo da Secretaria de Desenvolvimento Econômico de MG, em 2020 o ecossistema mineiro contava com mais de **1400** atores [1]. No entanto, esses levantamentos carecem de atualização contínua e de uma conexão direta com dados de produção científica. Ao criar uma plataforma que integra dados de diferentes fontes (acadêmicas e de mercado) por meio de técnicas de coleta automatizada de dados (*web scraping*), espera-se obter um panorama mais completo e atualizado da inovação no estado. Isso pode apoiar gestores públicos, investidores, universidades e empreendedores na tomada de decisões informadas, além de fomentar novas parcerias entre pesquisadores e empresas.

Objetivo Geral: O objetivo geral deste POC é desenvolver um mapeamento integrado do perfil de desenvolvimento e inovação em Minas Gerais, identificando as principais tecnologias e áreas de pesquisa em evidência no estado, mapeando startups e ICTs atuantes nesses domínios, e conectando esses elementos a pesquisadores locais por meio da coleta e integração de dados provenientes de diversas fontes (por exemplo, API do Semantic Scholar e bases de dados de startups). Em suma, busca-se criar uma base de conhecimento unificada que relacione atores e iniciativas do ecossistema de inovação mineiro.

Objetivos Específicos: Para atingir o objetivo geral proposto, foram delineados os seguintes objetivos específicos:

- Levantar e categorizar as principais tecnologias emergentes, áreas de pesquisa acadêmica, startups e ICTs no estado de Minas Gerais, a partir de fontes públicas e bases de dados disponíveis.
- Desenvolver procedimentos de coleta de dados automatizada (web scraping/API), incluindo o uso da API do Semantic Scholar para obter informações sobre publicações e pesquisadores, e fontes de dados sobre startups (por exemplo, diretórios de startups).
- Definir e implementar uma estratégia de armazenamento para os dados coletados, avaliando as alternativas entre banco de dados relacional, banco não relacional

(NoSQL) e data warehouse, de forma a escolher a solução mais adequada aos dados e consultas pretendidas.

- Integrar e correlacionar os dados coletados, estabelecendo ligações entre pesquisadores (e suas produções científicas) e os demais elementos do ecossistema de inovação (startups, tecnologias, ICTs), de modo a identificar convergências e possíveis parcerias ou lacunas.
- Elaborar visualizações, mapas ou relatórios analíticos que apresentem o panorama da inovação em Minas Gerais, evidenciando as conexões encontradas e fornecendo subsídios para conclusões e recomendações.

2 Referencial Teórico

Nesta seção são apresentados os conceitos e estudos que fundamentam o projeto, bem como uma análise de soluções existentes relacionadas ao mapeamento de ecossistemas de inovação, comparando-as à abordagem proposta neste TCC.

2.1 Conceitos Relevantes

Inovação tecnológica é definida de diversas formas na literatura, geralmente envolvendo a introdução de novos produtos, processos ou ideias que agregam valor econômico ou social. Trata-se de um processo dinâmico que envolve pesquisa, desenvolvimento e aplicação de conhecimentos [2]. No contexto regional, o conceito de **ecossistema de inovação** refere-se ao conjunto de atores interconectados (empresas, startups, universidades, centros de pesquisa, governo, investidores) cujas interações fomentam o empreendedorismo e a geração de novas tecnologias. Um componente importante desse ecossistema são as startups, empresas de base tecnológica caracterizadas pela busca de modelos de negócio inovadores e escaláveis. As ICTs (Instituições de Pesquisa Científica e Tecnológica), que incluem universidades, institutos de pesquisa e laboratórios, também desempenham um papel crucial ao gerar conhecimento e formar recursos humanos qualificados.

No caso de Minas Gerais, pode-se destacar a presença de um vigoroso **ecossistema de startups** especialmente na região metropolitana de Belo Horizonte (conhecida pelo apelido “San Pedro Valley”) e em outras cidades polo do estado. Tais startups frequentemente emergem em setores como tecnologia da informação, biotecnologia, mineração, agronegócio e energias renováveis. As interações entre startups e ICTs são frequentemente abordadas pelo modelo da *Tríplice Hélice*, que enfatiza a colaboração entre universidade, indústria e governo para impulsionar a inovação [2]. Nesse modelo, conectar pesquisadores acadêmicos com empreendedores e formuladores de políticas é essencial para transformar conhecimento em inovação prática.

2.2 Soluções Existentes na Literatura e no Mercado

A literatura registra diversos esforços de mapeamento e análise de ecossistemas de inovação regionais. Estudos acadêmicos frequentemente utilizam indicadores de inovação (como número de patentes, publicações científicas, startups fundadas, investimentos recebidos) para traçar comparativos entre regiões ou evoluções ao longo do tempo. Ferramentas de

visualização de redes já foram empregadas para entender colaborações entre universidades e empresas, evidenciando a formação de clusters tecnológicos.

No contexto de mercado e governo, iniciativas práticas de mapeamento também existem. Em Minas Gerais, por exemplo, o *Sistema Mineiro de Inovação (SIMI)* mantém desde 2016 uma base de dados do ecossistema inovador do estado, que inclui startups cadastradas, incubadoras, parques tecnológicos e outras entidades. A atualização de 2020 desse mapeamento [1] contabilizou centenas de startups e instituições em diversas regiões do estado, disponibilizando um panorama setorial e geográfico. Entretanto, essa plataforma depende de cadastro manual e não integra dados de produção científica ou detalhes aprofundados sobre as áreas de atuação dos pesquisadores.

No âmbito global, existem bases de dados como a *Crunchbase* e *AngelList* que reúnem informações sobre startups, e plataformas como a *OpenAlex* e *Semantic Scholar* que oferecem vastos dados sobre publicações e autores científicos. Contudo, a integração entre essas diferentes esferas de informação (acadêmica e de mercado) ainda é incipiente. Não há, até o momento, uma ferramenta específica focada em Minas Gerais que automatize a coleta e cruzamento de dados para revelar conexões entre pesquisadores e startups locais. Essa lacuna é justamente o espaço que este projeto pretende preencher.

2.3 Comparação e Contribuição da Proposta

Em comparação com as soluções existentes, a proposta deste TCC se diferencia por sua abordagem integradora e automatizada. Diferentemente de mapeamentos tradicionais que dependem de levantamento manual ou declaração voluntária de informações (como o SIMI ou diretórios de startups), aqui será empregada a técnica de *web scraping* e consumo de APIs para coletar dados de forma automática e abrangente. Isso permitirá uma atualização mais frequente e redução de vieses de seleção nos dados.

Além disso, ao conectar informações de pesquisas acadêmicas (por exemplo, publicações, keywords de pesquisa, currículos de pesquisadores) com dados de startups e empresas, espera-se revelar ligações que não são evidentes em cada fonte separadamente. Por exemplo, poderá ser possível identificar que determinada área tecnológica emergente (como inteligência artificial aplicada à saúde) possui simultaneamente grupos de pesquisa ativos em universidades mineiras e startups desenvolvendo produtos nesse setor, evidenciando um ponto forte regional. Inversamente, poderá-se detectar áreas em que há produção científica local significativa, mas poucas startups (ou vice-versa), indicando oportunidades de incentivo.

3 Metodologia

A metodologia do projeto abrange desde a coleta de informações até a análise e apresentação dos resultados, englobando também a escolha da infraestrutura de dados adequada. Os principais passos para a execução do projeto são descritos a seguir:

1. **Levantamento bibliográfico e planejamento:** Realizar uma revisão de literatura sobre ecossistemas de inovação, mapeamento de tecnologias e métodos de integração de dados. Refinar as perguntas de pesquisa e delinear os requisitos de informação (quais dados coletar e como).

2. **Identificação de fontes de dados:** Listar e avaliar possíveis fontes de informação relevantes. Isso incluirá APIs acadêmicas (como a do Semantic Scholar, e possivelmente OpenAlex), bases de dados de startups (por exemplo, *StartupBase* da Associação Brasileira de Startups, ou a base do SIMI), sites de instituições de pesquisa (universidades, FAPEMIG, CNPq etc.) e outras fontes públicas com dados de inovação.
3. **Desenvolvimento dos coletores de dados:** Implementar scripts e ferramentas de *web scraping* para extrair dados das fontes identificadas. Para fontes que oferecem API (como Semantic Scholar, CrossRef, etc.), desenvolver consultas para obter os dados desejados (por exemplo, buscar por autores afiliados a instituições mineiras ou por palavras-chave de tecnologias ligadas a Minas Gerais). Garantir que a coleta respeite políticas de uso e limites das APIs.
4. **Construção de base de dados:** desenvolvimento de uma estrutura de *Data Warehouse* ou repositório unificado para armazenar os dados coletados. Nessa fase, os dados de diferentes fontes serão integrados, garantindo consistência e eliminando duplicidades. Serão definidas dimensões de análise (por exemplo, por região geográfica, setor de atividade, tipo de ator) e realizadas transformações necessárias para viabilizar consultas e cruzamentos eficientes das informações.
5. **Análise e cruzamento de informações:** Com os dados integrados, executar consultas e análises para extrair insights. Por exemplo, gerar listas de quais tecnologias são mais mencionadas em pesquisas e simultaneamente focos de startups; mapear a rede de colaboração indicando se determinados pesquisadores têm ligação com startups (seja por participarem como fundadores/consultores ou coautoria com alguém da indústria); identificar regiões ou universidades destaque em certos domínios tecnológicos, etc. Poderão ser empregadas técnicas de análise de grafos e estatísticas descritivas.
6. **Elaboração dos produtos finais:** Preparar as formas de apresentação dos resultados: possivelmente um relatório escrito contendo tabelas e gráficos, mapas geográficos indicando concentração de startups e ICTs por região, e quiçá um protótipo de dashboard ou site interativo ilustrando o ecossistema.

Para a **coleta e análise de dados**, serão utilizadas principalmente ferramentas e linguagens de programação adequadas à manipulação de dados em larga escala. Espera-se utilizar Python para a implementação dos scripts de scraping e consumo de APIs (com bibliotecas como *requests*, *BeautifulSoup* para HTML, e bibliotecas específicas para APIs REST). Os dados coletados, que provavelmente estarão em formatos JSON, CSV ou HTML, serão então transformados e carregados na base de dados escolhida.

Durante a etapa de análise, ferramentas de visualização (como bibliotecas Python *matplotlib*, *seaborn* ou ferramentas de GIS para mapas) poderão ser utilizadas para gerar os gráficos e mapas desejados. A validação dos dados e resultados será feita comparando-se, quando possível, com estatísticas conhecidas (por exemplo, números publicados em relatórios oficiais como os do governo estadual ou associações de startups) de modo a verificar a consistência do que for encontrado.

4 Resultados Esperados

Ao final do projeto, espera-se obter diversos produtos e resultados de valor para a comunidade acadêmica e o ecossistema de inovação de Minas Gerais:

- **Base de dados integrada:** Um banco de dados consolidado contendo informações inter-relacionadas sobre pesquisadores (e suas produções), startups, tecnologias e ICTs de Minas Gerais. Essa base servirá como repositório central de conhecimento sobre a inovação no estado.
- **Mapeamento do ecossistema:** Representações visuais (mapas, gráficos e possivelmente redes) que ilustrem o panorama mapeado. Por exemplo, mapas temáticos mostrando a distribuição geográfica de startups e ICTs pelo estado, gráficos destacando as tecnologias mais presentes e redes que conectam pesquisadores a startups por meio de temas comuns.
- **Relatório analítico:** Um documento descrevendo os achados principais, incluindo a identificação de áreas tecnológicas em que Minas Gerais apresenta vantagem competitiva ou lacunas. Esse relatório incluirá interpretações dos dados coletados e poderá sugerir ações (como incentivos ou criação de programas de parceria) baseadas nas evidências.

Em suma, o trabalho deverá fornecer um retrato da convergência entre academia e mercado no cenário mineiro de inovação, além de um arcabouço de dados que pode continuar a ser alimentado e atualizado após a conclusão do POC, servindo de legado para pesquisas e iniciativas futuras.

5 Etapas e Cronograma

A execução do projeto está planejada para o período de maio a outubro de 2025, com cronograma semanal conforme a tabela a seguir. Cada atividade corresponde aos passos da metodologia e contempla tempo para imprevistos e ajustes:

Semana (2025)	Atividades Planejadas
Maio – Semana 1 (05-11/05)	Levantamento bibliográfico e refinamento das perguntas de pesquisa. Planejamento geral do projeto e definição dos dados e indicadores a serem coletados.
Maio – Semana 2 (12-18/05)	Continuação da revisão teórica. Identificação e análise de fontes de dados: APIs (Semantic Scholar, OpenAlex), diretórios de startups, bases governamentais.
Maio – Semana 3 (19-25/05)	Início da implementação dos coletores de dados (web scraping/API). Testes iniciais com a API do Semantic Scholar e coleta piloto.
Maio – Semana 4 (26/05-01/06)	Desenvolvimento contínuo dos scripts de coleta. Levantamento de dados sobre startups em MG e coleta de dados iniciais de pesquisadores.
Junho – Semana 5 (02-08/06)	Projeto do banco de dados: escolha entre SQL, NoSQL ou DW. Definição do modelo lógico da base (entidades e relacionamentos principais).
Junho – Semana 6 (09-15/06)	Criação da estrutura física da base de dados. Inserção dos primeiros dados limpos. Testes de integração e normalização.
Junho – Semana 7 (16-22/06)	Continuação da coleta e enriquecimento dos dados. Implementação de rotinas de transformação e padronização dos dados coletados.
Junho – Semana 8 (23-29/06)	Validação e verificação da qualidade dos dados. Resolução de problemas de duplicidade, inconsistências e erros de codificação.
Julho – Semana 9 (30/06-06/07)	Início da fase de análise: criação de consultas exploratórias e visualizações iniciais. Identificação de padrões e correlações.
Julho – Semana 10 (07-13/07)	Desenvolvimento de análises mais elaboradas: clusters, áreas tecnológicas em destaque, conexões entre startups e pesquisadores.
Julho – Semana 11 (14-20/07)	Geração de mapas e redes de interação. Análise geográfica e setorial do ecossistema de inovação em Minas Gerais.
Julho – Semana 12 (21-27/07)	Consolidação dos principais achados. Discussões iniciais com o orientador e ajustes no plano analítico.
Agosto – Semana 13 (28/07-03/08)	Redação da seção de Resultados. Integração entre análise empírica e fundamentação teórica. Inserção de tabelas, gráficos e mapas.
Agosto – Semana 14 (04-10/08)	Redação contínua dos resultados e início da seção de Discussão. Articulação dos achados com a literatura.
Agosto – Semana 15 (11-17/08)	Elaboração da seção de Conclusão: recomendações, limitações e possíveis desdobramentos futuros do projeto.
Agosto – Semana 16 (18-24/08)	Revisão geral do documento. Verificação de objetivos cumpridos, coerência textual e fluxo lógico entre seções.
Setembro – Semana 17 (25-31/08)	Preparação da versão preliminar do TCC. Padronização do documento, revisão gramatical e inserção de referências formatadas.
Setembro – Semana 18 (01-07/09)	Envio ao orientador. Recebimento de comentários e sugestões. Planejamento das revisões a serem aplicadas.

Este cronograma poderá ser ajustado conforme o andamento real do projeto, mas serve como guia para assegurar que todas as etapas sejam cumpridas dentro do prazo disponível.

6 Referências Bibliográficas

Referências

- [1] **SEDE/MG.** (2020). *Mapa do Ecossistema de Inovação Mineiro*. Secretaria de Estado de Desenvolvimento Econômico de Minas Gerais; SIMI & Liga Ventures, Belo Horizonte.
- [2] **Etzkowitz, H. & Leydesdorff, L.** (2000). The dynamics of innovation: from National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109-123.