

# Atividade 01 - Definição do tema

## Equipe

**Nome:** *TechDiscovers*

**Repositório GitLab da equipe:** <https://gitlab.com/bezenatti/tech-tendencies.git>

**Membros:**

- Bruno Emanuel Zenatti, 2264919, @bezenatti, brunozenatti@alunos.utfpr.edu.br, Engenharia da Computação, UTFPR
- Gabriel Cassol Bach, 2129035, @gabrielcassolbach, gabrielbach@alunos.utfpr.edu.br, Engenharia da Computação, UTFPR

## Tema

Mapeamento de tendências de tecnologia utilizando técnicas de PLN.

## Objetivos

- Identificar e Quantificar Tópicos Tecnológicos Emergentes
- Mapear a Evolução de Tecnologias Específicas
- Identificar os Principais Atores em Setores de Inovação
- Identificar itens que diferem do comportamento/conteúdo “normal”

## Fontes de dados

O Harvard USPTO Dataset (HUPD) é um corpus de grande escala, bem estruturado e multifuncional de pedidos de patentes de utilidade em língua inglesa, apresentados ao Escritório de Patentes e Marcas dos Estados Unidos (USPTO) entre janeiro de 2004 e dezembro de 2014 (<https://patentdataset.org/>).

## Desafios

- Linguagem específica e técnica
  - Para cada aplicação, têm diversos campos, neles incluem:
    - "abstract": "...",
    - "claims": "...",
    - "background": "...",
    - "summary": "...",
    - "full\_description": "..."
  - Os quais podem resultar em linguagem muito técnicas, nichadas e com muitas siglas.
- Termos ambíguos
- Identificação de informações relevantes