








# Software Analytics Canvas

**Project:** Extração de hard e soft-skills de descrições de vagas de trabalho.

 <h2>1. Question</h2> <p><i>What is it that we want to know about the software / processes / usage / organization / etc.?</i></p> <p>Extraír habilidades , técnicas e sociais, de descrições de vagas de emprego na área de tecnologia.</p>	 <h2>2. Data Sources</h2> <p><i>Which data can possibly answer our question? What information do we need?</i></p> <ul style="list-style-type: none"><li>As descrições de vagas de trabalho do setor de tecnologia foram retiradas do site <a href="#">Catho</a> e <a href="#">Remotar</a>. Se tratam de agregadores de vagas de trabalho, tendo a Remotar o foco de vagas com dinâmica remota.</li><li>Taxonomia de hard-skills extraída com o auxílio da <a href="#">EMSI</a></li><li>Taxonomia de soft-skills criada por anotação</li></ul>	 <h2>3. Heuristics</h2> <p><i>Which assumptions do we want to make to simplify the answer to our question?</i></p> <p>Vagas de trabalho na área de tecnologia possui, em média, uma melhor estrutura sendo mais clara perante as habilidades que são necessárias para ocupar a vaga. Sendo assim mais fácil de identificar os textos.</p>	 <h2>4. Validation</h2> <p><i>What results do we expect from our analysis, how are they reviewed and presented in an understandable way?</i></p> <p>Conseguir identificar/extrair as habilidades técnicas assim como habilidades sociais dentro das descrições.</p>
 <h2>5. Implementation</h2> <p><i>How can we implement the analysis step by step and in a comprehensible way?</i></p> <ul style="list-style-type: none"><li>Web scrap para extração dos textos das vagas de tecnologias nos sites</li><li>Tratamento de dados para texto</li><li>Criação de taxonomias para hard- skills e soft-skill</li><li>NER<ul style="list-style-type: none"><li>criação da base de treinamento</li></ul></li><li>Tokenização e Word embeddings<ul style="list-style-type: none"><li>Word2Vec, FastText e BERT</li></ul></li><li>Modelos de classificação múltipla:<ul style="list-style-type: none"><li>SVM</li></ul></li><li>Avaliação<ul style="list-style-type: none"><li>precision,recall</li><li>F1-score</li><li>Matriz de confusão</li></ul></li></ul>		 <h2>6. Results</h2> <p><i>What are the main insights from our analysis?</i></p>	
		 <h2>7. Next Steps</h2> <p><i>What follow-up actions can we derive from the findings? Who or what do we need to address next?</i></p>	

