

Domain Understanding Document

Project: Role2Skill

1. Domain Overview

This project is situated in the domain of the online job market, with a focus on IT and technical job vacancies. In this domain, job roles are described primarily through free-text job descriptions, which include responsibilities, requirements, and lists of desired skills.

The job market relies heavily on these textual descriptions to communicate expectations between employers and job seekers. As a result, the clarity and consistency of this text directly influence how accurately job roles can be interpreted and compared.

2. Domain Problem

A key problem within this domain is the lack of clear structure in how job roles are defined through skills. Job descriptions often contain long and inconsistent lists of skills, where important competencies are mixed with optional or vaguely defined requirements.

Because of this, it becomes difficult to determine which skills truly define a job role and how similar or different job roles are based on their required skills.

3. Domain Actors (Stakeholders)

Several actors interact within this domain. Employers create job vacancies and describe required skills in natural language. Job platforms host and distribute vacancy data. Job seekers, especially students and junior professionals, interpret these descriptions to guide their career preparation. Educational institutions use labor market information to align curricula with industry needs.

4. Core Domain Concepts

The domain is based on several key concepts. A job vacancy is a single job posting that describes an open position. A job role refers to the type of position defined by its title and required skills. A skill is a technical or soft ability mentioned in a job description. Skill representation means expressing these skills in a structured way. Skill overlap describes how different job roles share common skills, while role similarity shows how close job roles are to each other based on their skill sets.

5. Domain Data Characteristics

Data in this domain is text-heavy and unstructured. Job descriptions vary in terminology, length, and level of detail, and often contain inconsistent skill naming. These characteristics apply across job vacancy datasets in general.

6. Domain Challenges

Several challenges arise naturally from the domain itself. Skills are often described inconsistently across vacancies. Job titles do not always accurately reflect actual skill requirements. Similar roles may appear different due to wording rather than content. Platform-specific tendencies can influence which skills are highlighted.

7. Domain Boundaries

The domain of this project includes analysis of job roles through required skills, comparison of skill compositions across job vacancies, and identification of similarities between roles.

The domain explicitly excludes salary analysis or labor market forecasting, personalized career advice, resume or candidate profiling, and real-time job posting collection.

8. Domain Value

Understanding this domain allows job market data to be interpreted in a more structured way. By organizing skill information, job roles can be compared more clearly, which reduces ambiguity in job descriptions and helps focus on skills rather than job titles.