

Carson Crockett

Introduction

This report is an investigation into the possibility of using Ontology to develop a legacy system migration tool. The purpose of this tool is to allow both technical and non-technical employees to send queries about business rules and logic regarding customers, promotions, and deals embedded in a legacy system and get a response using information we gather from the legacy system, which has the information embedded within it. This report is investigating the possibility of using Ontology to look at both the data in production (migration) and data models and generate such a product.

What is Ontology?

“An Ontology is a set of concepts and categories in a subject area or domain that shows their properties and the relations between them” (1). Our proposed solution to this problem is to develop a RAG pipeline to repurpose LLaMa 3 to answer user provided queries in the context of relevant documentation provided and retrieved via the pipeline. This solution uses various methods to retrieve relevant documents from a knowledge base to provide as the context of the query. “[An Ontology forms] the structure on which knowledge is contextualized” (1).

Types of Ontology Knowledge Models

The simplest level of ontologies establish a continuous vocabulary. For example it would interpret synonyms such as client, customer, and purchases as client. The next level acts as a thesaurus to identify terms which relate to a concept. The next level is taxonomy which defines parent-child relationships, such as one product being a component of another (1). “An Ontology is a representation of the relationships of multiple taxonomies” (1). We can generate a knowledge graph to understand instances of relationships in the taxonomies, such as sales between entities (1).

Can we use an Ontology?

Use of an Ontology could be used in this project to structure the knowledge base we plan to use to relate customers to promotions and deals across the legacy system. I believe that to do this we would first need to extract and store the business logic from the COBOL system so that Ontology can then relate the existing customer documentation and relationships to the COBOL data. We could develop (or use an open-source, already developed) Ontology in the retrieval stage of our RAG pipeline. By using an Ontology to structure the relevant data in our knowledge base we can improve the pipeline’s ability to retrieve relevant information and provide a more personalized response for the user by having a better structured representation of the relationships of the data.

Developing an Ontology

Here I will summarize the approach to developing an Ontology described in the protégé Ontology development guide. Protégé is an open-source framework which can be used to develop an Ontology. There are three fundamental rules to developing an Ontology.

- 1). There is no correct way to model a domain. The best solution depends on the application
- 2) Ontology development is an iterative process

3) Concepts in the Ontology should be close to objects and relationships in your domain of interest. These are most likely nouns (objects) and verbs (relationships) in sentences that describe your domain.

There are many modeling approaches we can use when developing an Ontology. We need to determine which one is most appropriate. This is where the iterative process of design becomes important. The Ontology will need to be evaluated and debugged after the initial version is made. To do this we must have an understanding of the reality we are representing (2).

Sources:

<https://www.earley.com/insights/role-ontology-and-information-architecture-ai> (1)

https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (2)