International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

Use of OWL and Semantic Web Technologies at Pinterest

Pramita Kastha[1]

[1]*Department of Electronics and Communication Engineering, Indian Institute of Technology, Kharagpur*

Abstract – *Pinterest is a well-known website. It is a social bookmarking website which is used for finding ideas for cinema, art, interior designing etc. Recently, the company adopted Semantic Web Technologies and created what is known as the Pinterest Taste Graph. This is done to ease the representation of data on the site and also to help in ads targeting and content recommendation systems. In this paper, we mainly talk about the Pinterest taxonomy, which describes the engineering of the underlying OWL ontology, and is the backbone of the Pinterest Taste Graph.*

Key Words: **Knowledge Graph, OWL, Taxonomy, Pinterest, Ontology Engineering**

1. INTRODUCTION

Pinterest is a social bookmarking website which allows people to discover things they like or would be interested at present or in the future, for example windsurfing, places to travel such as beaches, mountains etc. Users, which in the site nomenclature are called "Pinners", bookmark web-content, called "Pins" to their boards. A Pin contains an image and a description of what that Pin is about. When we click the Pin, we get redirected to the website which is the source of the Pin.

To recommend relevant Pins to its users, Pinterest defines various interests on which it classifies the Pins. The interest follows a hierarchical structure, what is known as the "Pinterest Taxonomy", which is the subject matter in this paper.

2. NOMENCLATURE

We now formally define some common terms which is essential to discuss more on this topic:

Pin: An image with a description which contains a link to a related external URL, which can be bookmarked on Pinterest.

Pinner: A user on who on Pinterest.

Interest: A concept under which a Pin can be classified as.

Board: A collection of similar Pins, which a Pinner as Pinned.

Taxonomy: Hierarchical arrangement of interests.

Taste Graph: The knowledge graph of Pinterest.

3. ONTOLOGY, TAXONOMY, INTERESTS AND PINS

Initially, Pinterest used keyword-based understanding to model their content. They extracted keywords from each Pins, did clean ups, labelled the Pins with the provided keywords, and the after studying the users' engagement with certain types of Pins, classified the users with some keywords as well.

But recently, Pinterest added an interest-based ad targeting interface, which works by naming a set of related keywords as interests, and subsequently organized them in a tree structure, which then was seen by the advertisers, who would pick nodes from it for their ad campaigns.

Existing public taxonomies fall short of meeting the requirement; hence Pinterest has designed its own taxonomy, which is based on the kind of content it hosts, and is specifically focused on ads targeting.

Pinterest uses Web Ontology Language (OWL) to model their taxonomy, and WebProtege as the collaborative development environment.

4. REQUIREMENTS

Here, we list and discuss business and tooling requirements that the Pinterest team might have needed for the formation of the Taste Graph

4.1 Business Requirements

The Pinterest Taxonomy internally will categorize all the Pins and all the users currently on it, and externally to Power ads targeting. For this, the final knowledge representation needs to:

1. Be a single root tree structure, and not a directed acrylic graph (DAG).
2. Should allow addition of attributes (facets) to the interests. This is needed to support multiple perspectives of the categorization.
3. Should not contain ambiguous interests.
4. Should include interests that has a large number of Pins, and discard interests that has minimal or no pins at all.
5. All children of a particular parent must be mutually exclusive collectively exhaustive (MECE).

6. Quality is of the most importance. It needs to be given much more preference than the quantity

4.2 Tooling Requirements

To create the Pinterest taxonomy, Pinterest needs a tool that has:

1. The functionality for multiple editors which are located at different locations to work on a same project at the same time.
2. A functionality which allows one to track which interest in the taxonomy has been reviewed.
3. Ways to reorganize the taxonomy. We should be able to rename and interest, move an interest, deprecate and interest etc.
4. A good user interface (UI) and a great user experience (UX).
5. Ability to search for interests in taxonomy based on their annotations.
6. Ability to share links directly to content in the taxonomy.

5. DESIGN DECISIONS AND MODELING CHOICES

The following is a possible modeling choice and engineering

Convention on the Pinterest data:

Interest as Classes Ontologically, an instance of an interest represents a Pinner's particular, unique interest in something. An object of the class may be then a particular interest topic.

Interest Descriptions For each interest, we add a label plus synonyms and definitions. We record this information using the following annotation properties**:**

rdfs:label - the primary name/label for an interest that exists in a specific language.

skos:altLabel – for noting synonyms.

skos:definition – to include a 1-2 sentence textual definition which will serve to provide clarity to the meaning of any particular interest.

Naming conventions Use title cases names with spaces for interest names. Ontology engineering recommendations state that singular noun forms be used for entity names. However, based on Pinterest's top queries, a mix of singular and plural noun forms may be used.

Naming Ambiguity Some topic names which have different topics but the same names must clearly be separated or disambiguated.

6. DISCUSSION

During the development of Pinterest Taxonomy, the Pinterest team may have been face to face with some or all of the following challenges both with respect to tool support for ontology development, as well as in terms of modeling choices:

Multiple Inheritance and Cross-Vertical Interests Some interests tend to have multiple parents. For example, "Room Lighting" can come under "Room Décor", but it can also come under "Light Fixtures".

Multiple Relationship Types among Interests For example, an interest in "Heels" is a clear interest in "Shoes", however, "Diwali Decoration" will come under "Decoration", "Diwali Recipes" will come under "Recipe", again both will come under "Diwali".

New Global Classes and Association with Interests Users and pins are classified against interests, but we can also classify them against attributes. Thus, Attributes should be defined as global classes and should be associated with applicable interests via appropriate relationships.

7. CONCLUSIONS

The flexibility of the OWL ontology representation allows Pinterest Taxonomy to be easily expanded to a DAG. Adding further logical axiomatization to encode facets of interest offers possibility of improving downstream search and recommendation applications. Additionally, it is known that the new representation of the content has increased Pinterest's revenue gains.

REFERENCES

[1] Pinterest Engineering (2020, January 10) Interest Taxonomy: A knowledge graph management system for content understanding at Pinterest. Medium https://medium.com/pinterest-engineering/interest-taxonomy-a-knowledge-graph-management-system-for-content-understanding-at-pinterest-a6ae75c203fd.

[2] Gonçalves, Rafael & Horridge, Matthew & Li, Rui & Liu, Yu & Musen, Mark & Nyulas, Csongor & Obamos, Evelyn & Shrouty, Dhananjay & Temple, David. (2019). Use of OWL and Semantic Web Technologies at Pinterest.