1. Introduction

- Ontologies are typically monolithic, and knowledge therein is rich and covers multiple topics. To improve the reusability and interoperability of ontologies, a straightforward way is to extract ontology fragments that can behave in the same way as the original ontology in a specific context, but significantly smaller.

- **Modularization** [1] and **uniform interpolation (UI)** [2] are logic-based approaches developed for computing ontology fragments, both of which preserve all the logical consequences of a seed signature, which is a sub-ontology of the ontology.

- Nevertheless, very little attention has been paid to the problem of **term selection** (seed signature selection) for ontology extraction.

- We argue that both the **logical information** and **lexical information** (e.g., **annotations**) of an ontology can help in establishing the relevance between terms. Fig. 1 shows that the MentholSpray concept could be relevant to the Sports domain under certain circumstances (e.g., when answering queries regarding the treatment of an injury in a football match).

- In this work, we propose a novel term selection approach to discovering semantic relationships between two isolated groups of terms utilizing ontology embedding techniques.

2. The Proposed Method

- The signature sig(O) of an ontology O is the set of all concept names in O. Given an ontology O and a primitive seed signature \( \Sigma \subseteq \text{sig}(O) \) containing few concept names suggested by domain experts or simply selected by users, which are believed to be terms that can best summarize the topic of interest, our approach computes an extension \( \Sigma' \) of \( \Sigma \) in **three steps**.

- **Step 1** (concept representation learning): Transform all concept names \( A \in O \) into \( D \)-dimensional vectors using the OWL2Vec* [3] model, the computation of which is based on all metadata (including logical axioms and annotation axioms) of \( O \).

- **Step 2** (computing relevance value): The relevance value (ranged from 0 to 1, with 1 standing for the strongest relevance and 0 for the weakest relevance) is computed by a newly developed algorithm called Nearest Neighbor Ranking (NN-RANK), which is shown in Alg. 1.

- **Step 3** (relevance-based seed signature extension): Given a threshold \( \sigma \) at the scale of 0 to 1, the extended seed signature \( \Sigma' \) is generated by \( \Sigma' = \Sigma \cup \{ A \mid A \in \text{sig}(O) \wedge f(A, \Sigma) \geq \sigma \} \).

3. Empirical Evaluation on SNOMED CT

- In this experiment, our task was to predict concepts in SNOMED CT Refsets (which contain terms in a specific clinical domain) based on a seed signature (randomly or manually) selected from these Refsets.

- **Tab. 1** shows that the proposed method (NN-RANK) outperformed another logic-based term selection baselines, especially when fine-tuning is done on the concept embedding. Besides, NN-RANK slightly outperforms another embedding-based method.

- **NN-RANK** was designed to fit in the multi-clusters pattern of topic words in the embedding space (see Fig. 2), which made it more effective.

4. Case Study: Ontology Abstraction

- Moreover, we evaluated how modularization and uniform interpolation benefit from the seed signature extended by NN-RANK in the OWL ontology abstraction task on the HeLiS ontology (shown in Fig. 3).

- It is shown in Tab. 2 that, with the help of NN-RANK, both methods output fragments that cover more concepts relevant to the topic, while UI fragments cover fewer irrelevant concepts and have better compactness.

5. Conclusion

- This work made a preliminary attempt to address the problem of extending an given seed signature with new terms selected sophisticatedly through embedding-based computation of important metadata of an OWL ontology.

- The proposed method outperformed other baseline models in selecting terms relevant to a given seed signature on the SNOMED CT ontology, and proved to be helpful in enabling ontology abstraction techniques to cover more terms relevant to the topic.

6. References

