TECHNICAL REPORT

In this report, we present our main modelling and evaluation results using our developed framework.

1. We introduced an ontology pertaining to smart farming through the presentation of a detailed knowledge base centered on key environmental factors affecting crop yields. This is depicted in **Figure 1**, showcasing the hierarchical structure of our developed ontology, and in **Figure 2** illustrating the structure of the object properties.

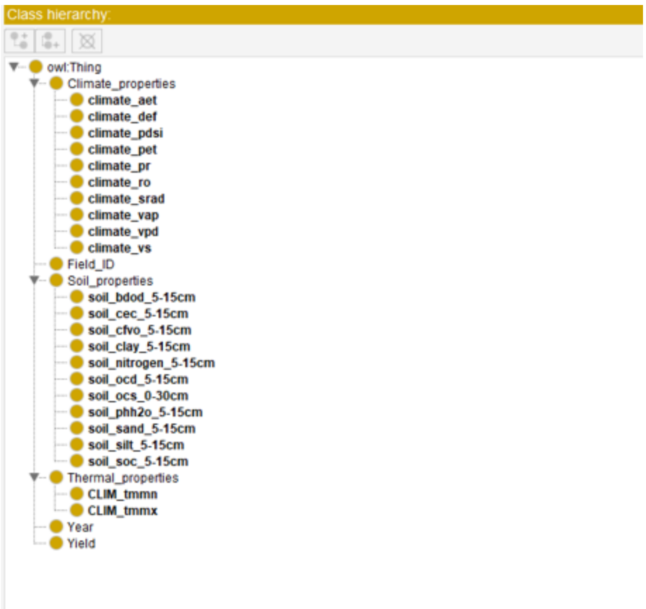


Figure 1. Hierarchical structure of our created ontology

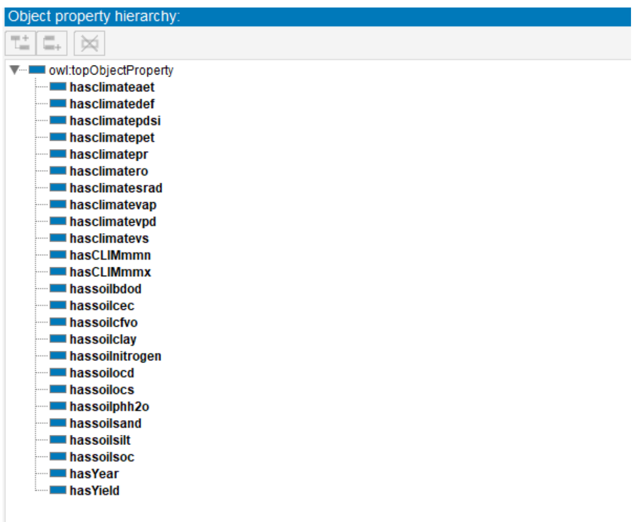


Figure 2. Object properties structure

2. We injected individuals into our ontology using transformation rules, as illustrated in **Figure 3**, utilizing data from CGIAR. This database contains environmental factors critical to agricultural yields, encompassing both climatic and soil characteristics. Soil attributes—such as Soil bdod, Soil cec, Soil cfvo, Soil clay, Soil nitrogen, Soil phh2o, Soil sand, Soil silt, Soil soc, Soil ocd, and Soil ocs—are stable factors significantly influencing crop production. These characteristics enable farmers to make well-informed decisions for various planting scenarios, thereby improving crop cultivation in ideal conditions and reducing losses under adverse circumstances. Climatic conditions form the second category in this database, representing dynamic factors that can substantially impact crop yields. Elements like droughts, water stress, and solar radiation are crucial, with the latter affecting the rate of photosynthesis on plant surfaces. Our ontology incorporates climatic properties such as Climate aet, Climate def, Climate pdsi, Climate pet, Climate pr, Climate ro, Climate srad, Climate vap, Climate vpd, and Climate vs, all of which have a direct correlation with crop production. Additionally, the ontology accounts for thermal properties, particularly focusing on the minimum and maximum temperature values.

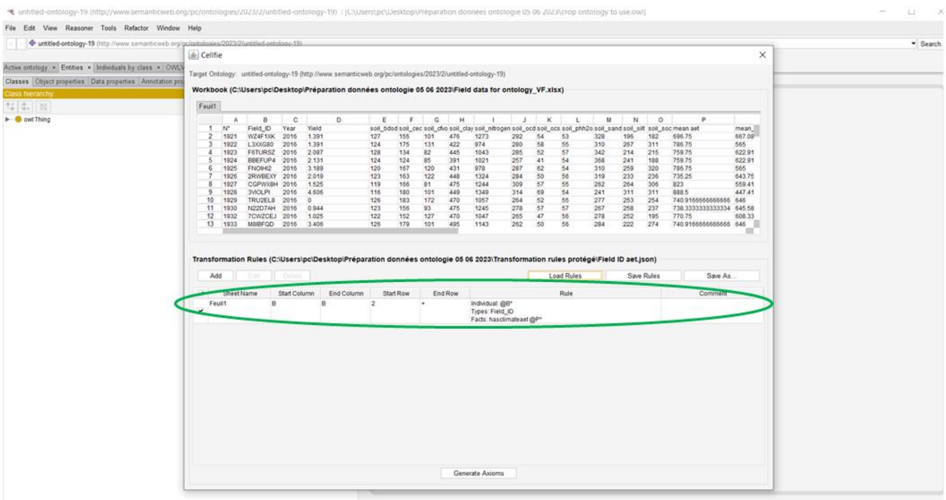
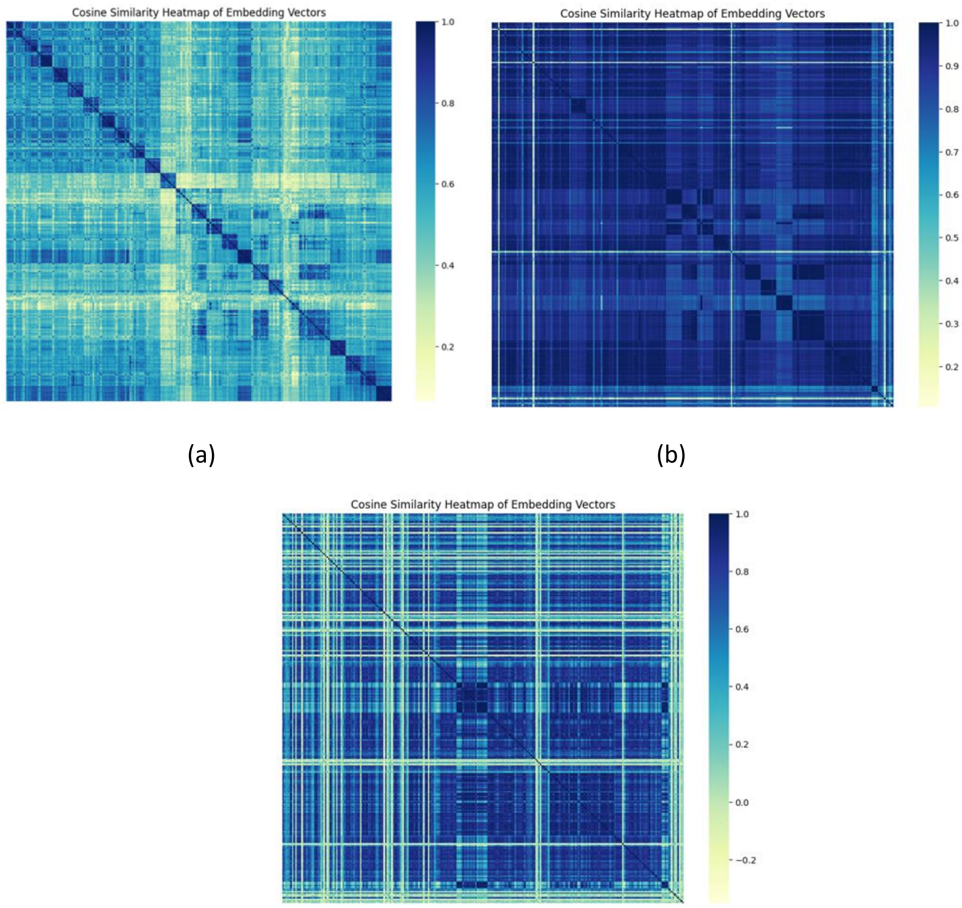


Figure 3. Ontology knowledge injection using transformation rules

3. Through the examination of the visual representations depicted in **Figure 4**, we can see that OWL2Vec* effectively distinguished between members of different classes, while also recognizing their underlying similarities. This is notably highlighted in the diagonal heatmap of **Figure 4-(a)**, where a more pronounced darker shade is observable. On the other hand, RDF2Vec produced vectors that were strikingly similar, even though they represented distinct properties. Meanwhile, Word2Vec, as illustrated in **Figure 4-(c)**, showed a moderate level of differentiation. Upon examination of the heatmaps, it becomes evident that OWL2Vec* exhibits consistently darker-colored pixels across the heatmap for each class, indicating higher cosine similarity measures among the ontology embedding vectors. This observation suggests that OWL2Vec* effectively captures semantic

relationships within the ontology. In contrast, the heatmaps generated for RDF2Vec and Word2Vec display a more varied distribution of pixel colors, with lighter shades interspersed, particularly for certain classes. This variance suggests less consistent similarity measures and potentially weaker semantic representations for RDF2Vec and Word2Vec in our scenario.



(a) (b) (c)

Figure 4. Cosine similarity heatmap using (a) OWL2Vec*, (b) RDF2Vec, and (c) Word2Vec

4. We also performed visualization through t-SNE in our quantitative framework, aiming to better compare the three embedding methods used in this study. By analyzing these representations illustrated in **Figures 5**, **6**, and **7**, it becomes apparent that OWL2Vec* effectively grouped similar instances and entities. In contrast, RDF2Vec struggled to differentiate between distinct classes, as evidenced by the nearly singular central cluster in the figure. On the other hand, Word2Vec exhibited only minor separations, leading to a more scattered visual depiction. Furthermore, upon closer inspection of the t-SNE plots of OWL2Vec*, specific clusters emerge, indicating cohesive groupings based on certain characteristics. For instance, the cluster representing the class 'Field ID' exhibits tightly grouped dots, suggesting instances with similar characteristics closely related to this class. Similarly, the cluster representing 'climate vpd' and 'soil phh2o' also demonstrates close proximity of dots, indicating shared features

among instances belonging to these classes. This analysis provides additional insight into the effectiveness of each embedding method in capturing and representing the underlying semantic structure of the ontology.

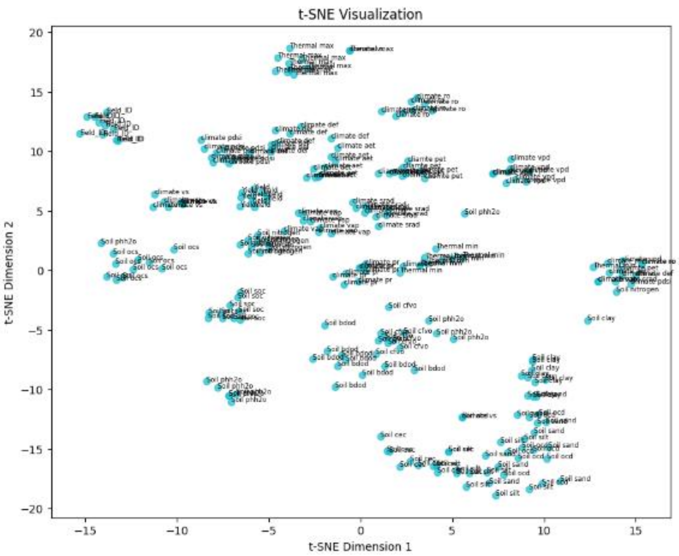


Figure 5. t-SNE visualization for OWL2Vec*

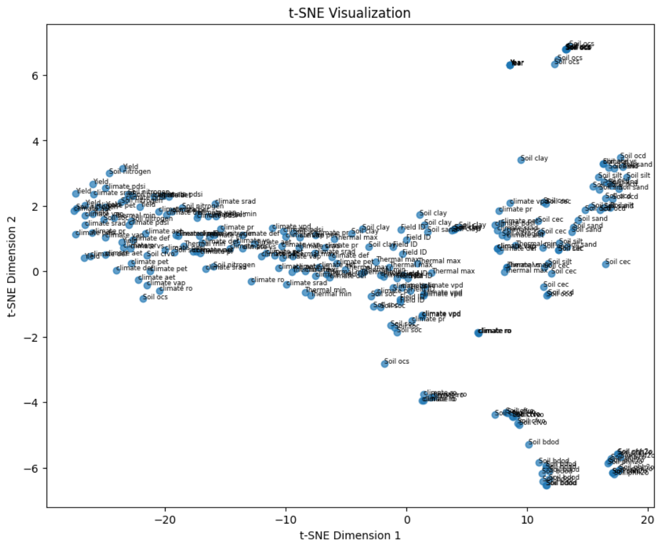


Figure 6. t-SNE visualization for RDF2Vec

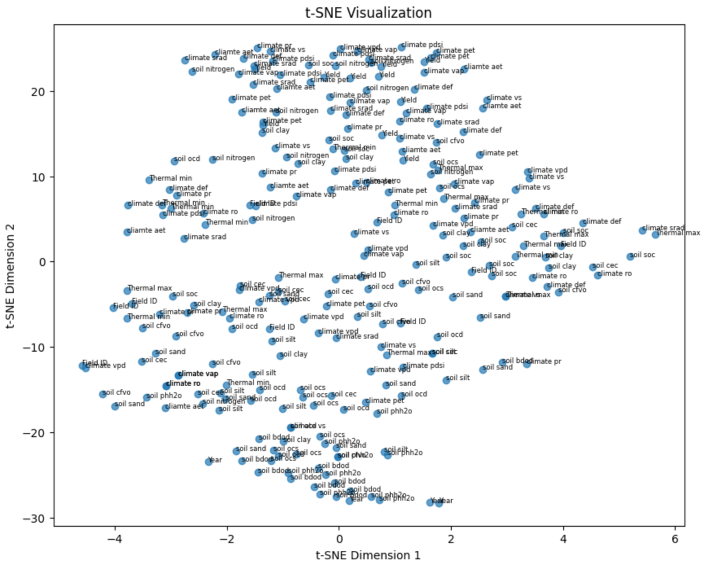


Figure 7. t-SNE visualization for Word2Vec