Practical Session 3: Web of Data - Ontology Alignment

Pablo Mollá Chárlez

Contents

1 Part 1: Ontology Alignment

Consider the ontology given in Figure 1, showing a set of classes organised by the relation **is-a**.

- **Question 1.** To compute the similarity between two ontologies by exploiting their structure, give the sets of classes that can be obtained by calculating the following relations on the ontology O:

$\boxed{\text{is-a}^{+}(\text{physician})}$ $\boxed{\text{is-a}^{-1}(\text{physician})}$ $\boxed{\text{is-a!}(\text{physician})}$

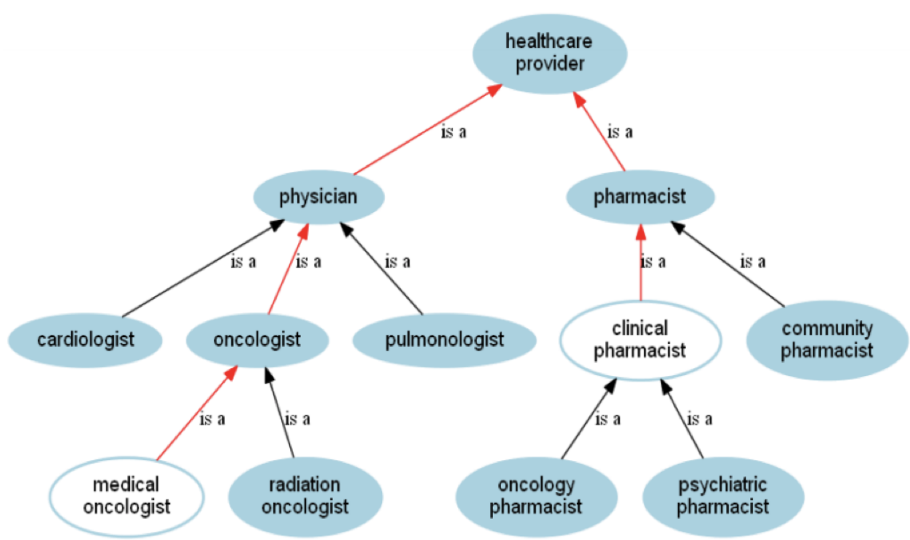

Figure 1: Medical Ontology 0

- **Question 2.** Given the ontologies O_1 and O_2 of figure 2, and the mapping O_1 : Artificial Intelligence $\equiv O_2$: AI, compute the Wu and Palmer scores for the following class pairs:
  1. (Machine Learning, Machine Translation)
  2. (Game Theory, Robotics)
  3. (Semantic Web, Semantics)

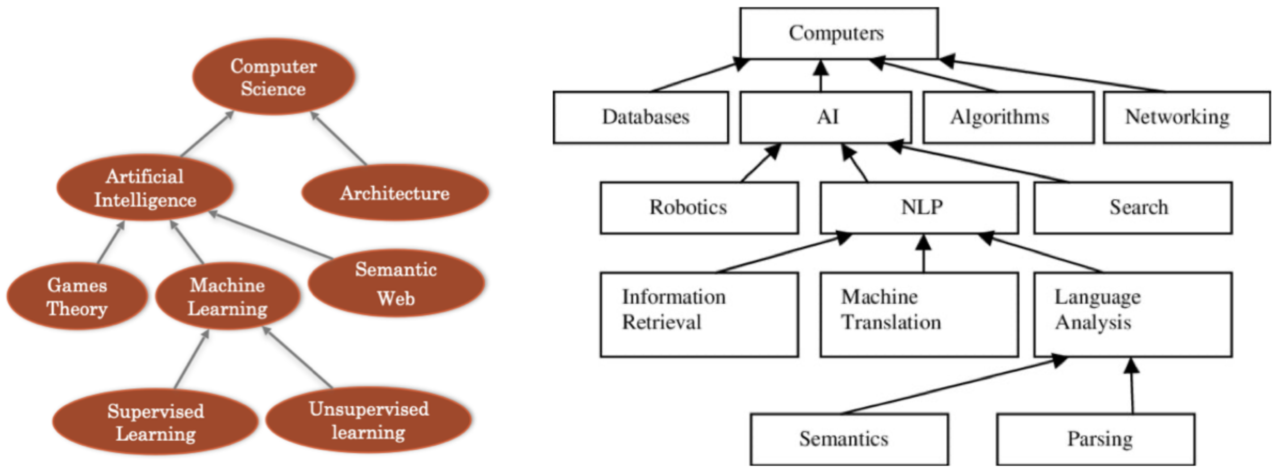

Figure 2: Ontologies O_1 and O_2 in Computer Science Field

- **Question 3.** For the same pairs of classes than in question 2, compute their similarity score using Jaccard measure. What are your comments on the results. If you compare the results of similarity computation that you obtained in question 2 and the one that you obtained in question 3, what is the best similarity measure to use for aligning these ontologies?

1.1 Answers

- **Question 1.** To compute the similarity between two ontologies by exploiting their structure, give the sets of classes that can be obtained by calculating the following relations on the ontology O:

  $\boxed{\text{is-a}^{+}(\text{physician})}$ $\boxed{\text{is-a}^{-1}(\text{physician})}$ $\boxed{\text{is-a!}(\text{physician})}$

  Let's first what it contains each set.

  - $is - a^{+}(physician)$: This set includes all classes that are descendants of the class physician in the ontology hierarchy. It traverses the **is-a** relation downward, including the class itself and all its subclasses.

  - $is - a^{-1}(physician)$: This set includes all classes that are ancestors of the class physician in the ontology hierarchy. It traverses the **is-a** relation upward, including the class itself and all its superclasses.

  - $is - a^{!}(physician)$: This set includes all classes that are siblings of the class physician in the ontology. Siblings share the same immediate parent in the **is-a** hierarchy, and does not descend into their subclasses. It excludes any descendants of those siblings.

  The sets of classes previously mentioned are the following:

  - $\boxed{\text{is-}a\text{+(physician)}}$ = {physician, cardiologist, oncologist, pulmonologist, medical oncologist, radiation oncologist}

  - $\boxed{\text{is-}a^{-1}\text{(physician)}}$ = {physician, healthcare provider}

  - $\boxed{\text{is-}a\text{!(physician)}}$ = {cardiologist, oncologist, pulmonologist}

- **Question 2.** Given the ontologies O_1 and O_2 of figure 2, and the mapping O_1 : Artificial Intelligence $\equiv O_2$: AI, compute the Wu and Palmer scores for the following class pairs:

  The **Wu and Palmer (WUP) Similarity Measure**, introduced by Wu and Palmer in 1994, is a semantic similarity metric used to quantify how similar two concepts (C_1 and C_2) are within a hierarchical ontology, such as WordNet. This measure is particularly useful in **ontology alignment** and **knowledge graph** applications to assess the relatedness of different entities.

$$\mathrm{Sim}(C_1, C_2) = \frac{2 \cdot \mathrm{level}(C)}{\mathrm{level}(C_1) + \mathrm{level}(C_2)}$$

  Where the components are:

  - **Level(C)**: Represents the depth of the **Least Common Subsumer (LCS)** in the ontology's hierarchy, an is defined as the most specific concept that is an ancestor of both C_1 and C_2.
  - **Level(C_1)** and **Level(C_2)**: Indicate the depth of each individual concept (C_1 and C_2) from the root of the ontology.

  The **Wu and Palmer similarity scores** are:

  - $Sim(\text{Machine Learning}, \text{Machine Translation}) = \frac{2 \cdot \text{Level(Artificial Intelligence)}}{\text{Level(Machine Learning)} + \text{Level(Machine Translation)}} = \frac{2 \cdot 2}{3+4} = \frac{4}{7}$
  - $Sim(\text{Game Theory}, \text{Robotics}) = \frac{2 \cdot \text{Level(Artificial Intelligence)}}{\text{Level(Game Theory)} + \text{Level(Robotics)}} = \frac{2 \cdot 2}{3+3} = \frac{4}{6} = \frac{2}{3}$
  - $Sim(\text{Semantic Web}, \text{Semantics}) = \frac{2 \cdot \text{Level(Artificial Intelligence)}}{\text{Level(Semantic Web)} + \text{Level(Semantics)}} = \frac{2 \cdot 2}{3+5} = \frac{4}{8} = \frac{1}{2}$

- **Question 3.** For the same pairs of classes than in question 2, compute their similarity score using Jaccard measure. What are your comments on the results. If you compare the results of similarity computation that you obtained in question 2 and the one that you obtained in question 3, what is the best similarity measure to use for aligning these ontologies?

  The Jaccard Similarity measures the overlap between two sets of tokens relative to their union and is defined as:

$$\mathrm{Jaccard}(S, T) = \frac{|S \cap T|}{|S \cup T|}$$

  Therefore, in our scenario, the Jaccard similarity scores are:

  - $Jaccard(\text{Machine Learning}, \text{Machine Translation}) = \frac{|\{'Machine', 'Learning'\} \cap \{'Machine', 'Translation'\}|}{|\{'Machine', 'Learning'\} \cup \{'Machine', 'Translation'\}|} = \frac{|\{'Machine'\}|}{|\{'Machine', 'Learning', 'Translation'\}|} = \frac{1}{3}$
  - $Jaccard(\text{Game Theory}, \text{Robotics}) = \frac{|\{'Game', 'Theory'\} \cap \{'Robotics'\}|}{|\{'Game', 'Theory'\} \cup \{'Robotics'\}|} = \frac{|\varnothing|}{|\{'Game', 'Theory', 'Robotics'\}|} = 0$
  - $Jaccard(\text{Semantic Web}, \text{Semantics}) = \frac{|\{'Semantic', 'Web'\} \cap \{'Semantics'\}|}{|\{'Semantic', 'Web'\} \cup \{'Semantics'\}|} = \frac{|\varnothing|}{|\{'Semantic', 'Web', 'Semantics'\}|} = 0$

  As we can observe, we obtain different scores with both similarity measures, and we can notice that Wu and Palmer is less strict as it produces higher scores. Therefore, we can't really tell which similarity measure is more adequate for aligning these 2 ontologies O_1 and O_2.