CSI5180 Topics in AI - Ontologies and Semantic Web
Winter 2019

Assignment 4 - Ontology Learning

Objective	Become familiar with ontology learning through textual analysis. **For this assignment, you have a choice between "hands-on" or "critical literature review" work.**
Due date	April 9th 2019, 11:30pm.
What to submit ?	Hands-On option Submit a report, in pdf format, which will include a title page (name, student number) as well as answers as specified for each of the questions below. Also submit the code you programmed. Make the report self-contained, by which I mean that I should be able to read it without having to go back to the assignment description. Critical literature review option Submit a report, in pdf format, see description later in this document.
How to submit ?	In Brightspace, through the link provided under the Assignment module.
Penalty	-10% per day late
Percentage	15% of overall semester grade
Software Requirement	If you choose the hands-on option, you can program in Java/Python, as you wish.

Question 1 - Corpus Building (5 points)

The first step of acquiring knowledge from text to build or expand an ontology is to find good sources of information. Go back to the ontology you developed in Assignment 2, and find 4-5 good documents for that topic. Construct a corpus from those documents.

IN YOUR REPORT: Describe the process you followed to find the sources of information. Why do you consider the documents "good". Provide the sources.

Question 2 - Term Extraction (10 points)

a) The second step is to find the terms in the corpus. Test a couple term extractors, or program your own.

b) As you already had performed the process manually in Assignment 2, you can now compare the result of the term extractors with your list of instances, classes and properties that you had defined. What is similar, what is missing?

IN YOUR REPORT: Describe your term extraction process. Provide the sources of the extractors used, or the source of the algorithm that you programmed. Show the results obtained. Discuss the comparison between your manually built ontology and your extraction process.

Question 3 - Taxonomy induction (10 points)

We saw two approaches in class to induce a taxonomy.

(a) Pattern-Based Approach:

(i) in your corpus, search for patterns (e.g. "X is a kind of Y", "Y such as X1, X2"). This is usually performed in a bootstrapping manner, but here, you can simply manually pre-program 3-4 patterns and search for them. Does that give interesting results? If yes, why? If no, why?
(ii) Also, try to search for superclass-subclass pairs from your ontology in Assignment 2. For example, if your ontology contained "baseball" subclass of "sport", search for sentences in your corpus that contain both words. Do these sentences contain interesting taxonomy patterns?? Explore.

(b) Term analysis (bracketing) approach: Start from the extracted terms and use some strategies to split/compare longer strings to shorter ones (e.g. "color laser printer" as subclass of both "color printer" and "laser printer", or "ice cream truck" to be a subclass of "truck" -- not "ice truck" and "cream truck"!!). You can define your own strategy, or be inspired from what you find in the literature, or what we explored in class.

Choose either (a) or (b) and implement it. Discuss the obtained results. Are they good? Does the taxonomy obtained resemble the taxonomy you did by hand in Assignment 2?

IN YOUR REPORT: Describe your taxonomy induction approach. Discuss the results. Are these results similar to what you have in your manually done ontology?

CRITICAL LITERATURE REVIEW OPTION

There is a large body of literature on various approaches (other than pattern-based as we've seen in class) to perform taxonomy induction.

Find 4-5 recent (in the past couple years) research articles of interest, read, summarize and compare them. Try to find articles presenting different approaches. Often the same authors publish multiple papers about one idea, please avoid gathering papers from the same authors unless they really present different ideas.

IN YOUR REPORT:

1. Title Page
2. Provide a title, a summary (in your own words) of each selected article, as well as a reason for selecting that article (why did you choose it ?)
3. Provide a critical view of each article (positive points, negative points)
4. Provide a comparative table in which you establish a few comparison points, and provide a comparative view of the articles.
5. Provide some examples from your ontology in Assignment 2 that seem to "fit" the algorithms. By which I mean, now that you've done a taxonomy manually, do you have any insight as to which algorithm is likely to work best for your topic.
6. Discussion, Conclusion
7. References