**Experimenting with EMAP, GO, FMA and CHEBI Ontologies**

This supplementary document explains regression model generations with ontologies other than SNOMED CT. We preferred the ontology SNOMED CT to present examples in the paper using, because we have observed in our experiments that:

1. SNOMED CT is the only ontology which gave OOM,
2. SNOMED CT has reached maximum axiom count in our experiments, mainly because its bigger size when compared to other ontologies,
3. SNOMED CT has a more variable graphic of observations which makes model generation more difficult when compared to other ontologies which generally provide a more linear behaviour.

To prove the generalizability of the proposed model, we would present more experiments with more ontologies. Because of the space limitations, we presented the results with SNOMED CT on the paper and presented 4 other ontologies which are widely used in testing algorithms/reasoners for high resource-consumption cases.

The folder of this document contains all processed steps for reaching the conclusion proposed here.

**1. Obtaining Ontologies**

Experimental ontologies used in document can be downloaded from these links:

EMAP: <http://purl.obolibrary.org/obo/emap.owl>

GO: <http://purl.obolibrary.org/obo/go.owl>

FMA-el (v.3.2.1): <https://storage.googleapis.com/google-code-archive-downloads/v2/code.google.com/fma-in-owl/fma-el.owl>

CHEBI: <http://purl.obolibrary.org/obo/chebi.owl>

**2. Pre-porocessing Ontologies**

As our experimental reasoners both support FunctionalSyntax format for input ontologies, all ontologies are translated to FunctionalSyntax by using “Translator.java” class in <https://github.com/IsaGuclu/KnowledgeExtraction>

During this translation *Annotation axioms are discarded* as ELK reasoner discard these axioms during classification.

**3. Knowledge Extraction via Agnostic Strategy**

Ontologies are extracted using different “desired signature”s and different axiom counts. For example, “chebi.owl” is extracted from 1,000 axioms to 350,000 as the annotation-discarded functional version contains 340,301 axioms. This extraction is done by using “ExtractorExtended.java” class in <https://github.com/IsaGuclu/KnowledgeExtraction> .

All needed ontologies (e.g., chebi.owl), axiom counts parameters (e.g., AxiomCounts\_chebi.txt) and designed signatures parameters (e.g., Desired\_1sp\_chebi.txt) are provided in <https://github.com/IsaGuclu/KnowledgeExtraction>

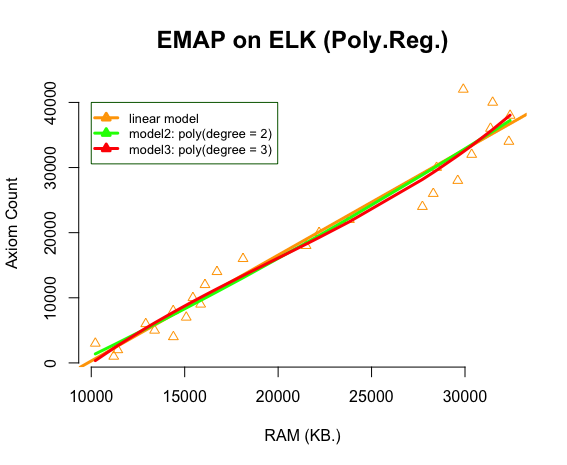
**4. Model Generation**

Extracted Ontologies are processed in Machine1 (?ref?) with obtaining max RAM amount before process finishes. After execution logs are collected for different runs, they are used for regression model generation.

R-scripts used in these model generation are also included in the subfolders of this repository. For generating an optimum regression model, a linear model is tested for every case, then a polynomial model with degree=2, then with degree=3 and degree=4.

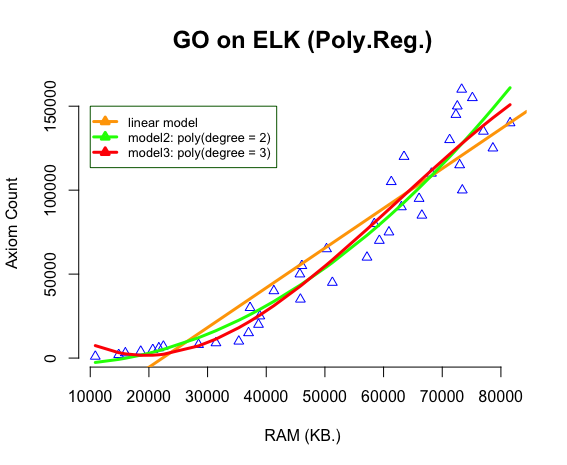
As we have observed misleading prediction results with degree=4 in extrapolation cases and degree=3 gave better results, **especially for high RAM consumption cases**, than previous models we hypothesised a polynomial model with degree=3 for these scenarios and in the paper.

**a. Model Generation via EMAP Ontology**



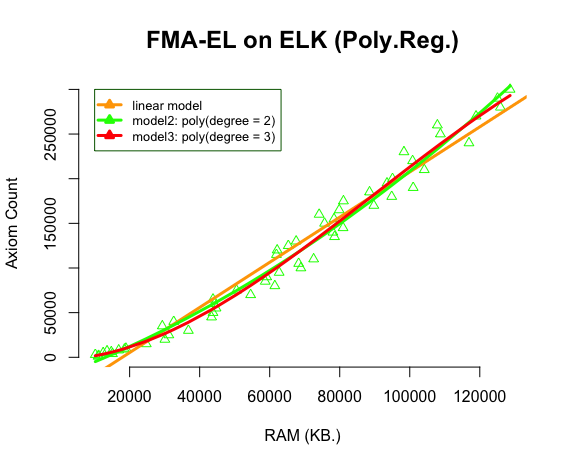
As seen in the diagram above, models generated with linear regression or polynomial regression with degree=2 and degree=3 don’t show a big difference. We claim that one reason for this result can be small axiom count of EMAP ontology (41,177).

**b. Model Generation via GO Ontology**



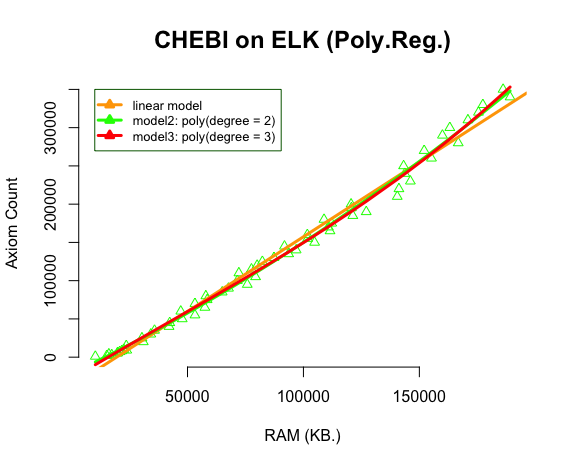
As seen in the diagram above, models generated with linear regression or polynomial regression with degree=2 and degree=3 show a little difference when compared to EMAP case. We claim that one reason for this result can axiom count of GO ontology (156,429).

**c. Model Generation via FMA-el Ontology**



As seen in the diagram above, models generated with linear regression or polynomial regression with degree=2 and degree=3 again show a little difference when compared to previous cases. FMA-el ontology contains 295,729 axioms.

**d. Model Generation via CHEBI Ontology**



As seen in the diagram above, models generated with linear regression or polynomial regression with degree=2 and degree=3 again show a little difference when compared to previous cases. CHEBI ontology contains 340,301 axioms.

**5. Conclusion**

In our experiments, we have observed that there is a significant relation between axiom count and RAM consumption using the ontologies stated in this document, i.e., EMAP, GO, FMA and CHEBI.

To present the accuracy of the proposed hypothesis and model in the paper, we preferred to continue experiments on SNOMED CT (Jan.2017), which shows a varying behaviour, more axiom counts and OOM when threshold is crossed.