→ SYSTEM ENGINEERING MODELS MEET KNOWLEDGE GRAPHS

Executive Report Summary

ESA Contract No. 4000133311/20/NL/GLC

Paul Darm

Audrey Berquand

Annalisa Riccardi

Edmondo Minisci

European Space Agency
European Space Research and Technology Centre
Keplerlaan 1, 2201 AZ Noordwijk

June, 2022

1 EXECUTIVE SUMMARY

1.1 Introduction

Tables and graphs are knowingly used to organise data within a company with different levels of depth and complexity. Knowledge graphs (KGs) are particularly useful because they can cope with data diversity (high-quality complete data and sparse and incomplete data), they have a high degree of scalability and flexibility (the semantic data model can be inter-operational, large, wide and as deep as needed) and, last but not least, they provide reasoning and inference capabilities. On the other hand, the need for intelligent systems enabling the access to heritage information is becoming more and more pressing with the growing amount of accumulated data. This is especially relevant for Model Based System Engineering (MBSE) design where the creation of new complex concepts is facilitated by agreed data standards and data synchronisation tools. In this activity, ESA Engineering Models (EMs) based on the data model defined in the ECSS-E-TM-10-25A Annex A are migrated to a Knowledge Graph. The graph is further enriched with metadata information collected from the mission's feasibility reports. The resulting graph is first used to investigate similarities between past missions and to identify common subsystems architectures. Then, based on the knowledge contained in the KG, a novel recommender system to suggest engineering components for new missions is deployed.

1.2 Similarity metrics

Similarity is computed automatically by the system between mission metadata entries and EMs components.

Mission metadata entries are defined as a list of numerical or textual information. These are for example, orbit type, mission objectives, requirements, ... Numerical metadata can be easily compared directly by using their values. However most of the metadata parameters take free text as input though, and thus Natural Language Processing has to be used to transform the text into a format for similarity computation. The current state of the art for comparing textual information is to use artificial neural networks, which transform natural language into a vector representation, usually referred to as "embedding". The similarity between the vectors can then be calculated for instance with the cosine similarity metric.

Besides comparing the EMs based on the available metadata information for the associated mission, the design information of the EMs can also be used directly. One approach is to compare the -nents included in the different EMs with each other. Clusters of same or similar components could reveal patterns in the design for specific mission types. Each component has specific associated parameters. Based on sharing parameters, one can establish how similar two elements are to each other. The Jaccard similarity index compares the sets of parameters of two components to see which parameters are shared and which are distinct. It's a measure of similarity for the two sets of parameters, with a range from 0 to 1. The higher the value, the more similar the two sets are.

1.3 Recommendation System

The scenario is to compare, for a known set of metadata of a future mission, the contents of the KG, identifying past similar studies and looking for patterns of common components in their associated design architecture. These common components can then be used as a form of recommendation, for "kick-starting" the design of a new mission. Firstly, the cosine similarities between the metadata of past studies in the KG and the metadata of a new study are calculated and stored. Then, these calculated similarities values are used in a clustering process, identifying relevant past missions in respect to the new mission. Depending to which cluster of missions the metadata information of the new mission belongs to, the components of the EMs in that cluster are then investigated for similar components. Based on either the Jaccard or cosine similarity between the elements, accumulations of similar elements in the cluster of missions can be used as possible recommendations for components in the future mission. In Figure 1.1 a schematic representation of the whole system architecture is reported where the main components are

- **Data (EMs Metadata, EMs)**: heritage information fed into the system to provide recommendations, these are past mission studies and corresponding mission metadata (ex: mission objectives, orbit type, ...)
- **Knowledge Graph (KG)**: graph structured database that merges in a single data structure all the heritage information about past missions
- **Computational Engines**:
  - Similarities: set of routines to compute similarity scores between metadata and system components
  - Clustering: set of routines to cluster missions based on a overall similarity score or metadata subsets similarity scores
- **Recommendations**: set of routines that by ranking and sorting the results of the similarity analysis provide components recommendations to the "New Mission" metadata fed as input to the system

1.4 Results and Conclusions

KGs have been identified as suitable tools for firstly providing a framework for implementation of multiple EMs and secondly for a further analysis of their contents. Afterwards, a methodology was introduced to firstly compare EMs based on their mission specific associated metadata, and secondly to identify common components in their EMs. During the study it was demonstrated that for a limited selection of missions the KG can be used to convincingly identify for some metadata parameters common components in past-mission EMs for future mission design. Based on the mission description, it was demonstrated for a new X-ray observatory mission, that two past missions could be identified also concerned about X-ray sources in the universe. This enabled the comparison of similar components where for this metadata parameter and combination of similar past mission studies 83 out of 114 possible elements were identified from the closest mission as a recommendation. Nevertheless, further testing for more different new mission concepts and for additional meta- data parameters is needed to spot possible weak points in the analysis. For instance in the analysis for a weighted similarity of the defined metadata parameters in the case study, missions were identified as similar to the new mission, which under closer inspection were actually only loosely related. Therefore, the methodology is still requiring expert knowledge in order to analyse and validate the results from the

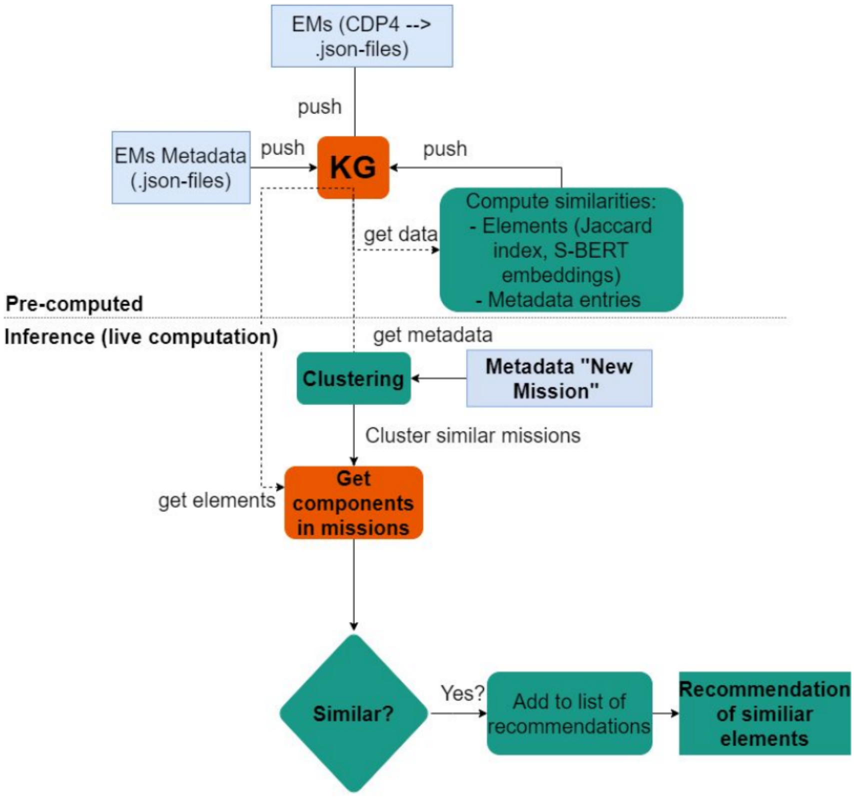

Figure 1.1: Recommendation System - architecture

similar missions clusters as well as component similarity analysis. Possible lessons learned, showcased the demand and potential of further applications of NLP in order to provide a more thorough and detailed description of the elements in the KG, e.g. linking the information found in the study reports automatically to the entities found in the KG. The concept of a recommendation system for spacecraft design could be expanded in the future to include a wider range of data from the EMs. Right now the recommendation is limited on analysis of the static representation of metadata and elements in the last iteration of a mission. In the future, the evolution of a mission concept overtime could also be used for potential insight, e.g. how metadata parameters are changing over different design iterations or even how design iterations change by themselves. One potentially insight could be what kind of elements or parameter updates usually happen in conjunction. This could provide another source for expert knowledge and thus support new spacecraft design.