

52nd IAA SYMPOSIUM ON SAFETY, QUALITY AND KNOWLEDGE MANAGEMENT IN SPACE
ACTIVITIES (D5)

Knowledge management for space activities in the digital era (2)

Author: Ms. Audrey Berquand

University of Strathclyde, United Kingdom, audrey.berquand@strath.ac.uk

Mr. Iain McDonald

University of Strathclyde, United Kingdom, iain.mcdonald.2015@uni.strath.ac.uk

Dr. Annalisa Riccardi

University of Strathclyde, United Kingdom, annalisa.riccardi@strath.ac.uk

Dr. Yashar Moshfeghi

University of Strathclyde, United Kingdom, yashar.moshfeghi@strath.ac.uk

THE AUTOMATIC CATEGORISATION OF SPACE MISSION REQUIREMENTS FOR THE DESIGN
ENGINEERING ASSISTANT**Abstract**

The Design Engineering Assistant (DEA) is an expert system for the early stages of space mission design. The DEA will provide an easy and fast access to accumulated unstructured (textual data) and semi-structured data from the field of space mission design. The experts will be able to query the DEA's knowledge base via a User Interface. The Artificial Intelligence (AI) behind the DEA has to not only be able to query the DEA knowledge base when prompted, but also grasp the context or the meaning behind the User query so that the most relevant knowledge can be extracted.

To perform more efficient querying, the generally unstructured data stored from previous space missions' feasibility studies reports has to be organised into a structured format. One method to accomplish this is through the use of topic modelling. Topic modelling allows for a document to be interpreted as containing a predefined number of topics, with each topic defined by a dictionary of related words, also called "Bags of Words". Using a corpus of documents, a model containing a variety of different topics can be generated and then used to categorise the report content. The method of topic modelling chosen for this study is Latent Dirichlet Allocation (LDA), which relies on a probability based approach to generate the words belonging to a topic.

This paper will show how the dictionaries of topics used to categorise input requirements can be automatically extracted from a corpus of unstructured data. A list of subfields related to space mission design such as propulsion, structures or thermal has been initially defined. The unstructured data is then associated to one or more subfields based on how the pages content matches with the topic dictionaries. As the dictionaries are generated from Wikipedia pages, it should allow for non-technical definitions of the topics, and therefore be more robust.

The results of testing the model will be presented as various sample requirement queries being submitted and shown, to match with the most relevant topics, something a human expert can do with ease. In the completed DEA, once a User query is submitted, the User should then be directed to the correct subfield of design. Through the use of word weightings being assigned in the LDA process, the likelihood of which subsections are relevant will also be shown, emulating the relevance of the query outputs.