**Similarity Measurement for Ontology-Based Product Configurations**

N. Baumbach*1; D. Peters1\*, L. Wagner2, S. Humbla 2*

*1DLR Institute of Data Science, Jena, Germany,*

*2Jena-Optronik, Jena, Germany,*

*\*Corresponding author contact details:* [*diana.peters@dlr.de*](mailto:diana.peters@dlr.de)

In several technical domains, suppliers describe the technical details of their products in data sheets. In domains like space industry where products are built upon request, tailored to customer needs and not sold off the shelf, data sheets describe baselines and production capabilities rather than fixed products. Therefore, one data sheet describes often a whole product family and one specific configuration leads to a single product.

Every new configuration needs to pass several steps of verification and testing to make sure it meets the required standards, for example ECSS. Especially testing can be very expensive and time consuming. A lot of configurations are only built a handful of times, which increases the impact of testing on overall cost even more. In some cases, a similar, already tested product of slight different configuration would equally be suited.

In this paper, we present the application of different similarity measures to product configurations in the space domain. We evaluated the similarity measures using the star tracker ASTRO APS of German manufacturer Jena-Optronik GmbH as an example. The formal knowledge about possible configurations is stored in an ontology.

We evaluated three similarity measures with differing complexity: the Jaccard index based on comparison of set intersections, a measure proposed by Ahmand and Dey for their k-mean clustering algorithm for mixed data, and a new measure, developed by us, that recursively compares graph representations of configurations.

Every similarity measure was applied to a set of test cases and the results were evaluated by domain experts of Jena-Optronik GmbH. We rank the similarity measures according to their performance in finding similar product configurations.

**References**

[1] Paul Jaccard. „Lois de distribution florale dans la zone alpine“. In: Bulletin

de la Société vaudoise des sciences naturelles 38 (Jan. 1902), S. 69–130.

[2] Amir Ahmad und Lipika Dey. „A k-mean clustering algorithm for mixed

numeric and categorical data“. In: Data & Knowledge Engineering 63.2 (2007),

S. 503–527. ISSN: 0169-023X.