Chapter 10
Generate and Test for Formulated Product Variants With Information Extraction and an In-Silico Model

Sagar Sunkle
TCS Research, Tata Consultancy Services, India

Deepak Jain
https://orcid.org/0000-0001-5849-454X
TCS Research, Tata Consultancy Services, India

Krati Saxena
TCS Research, Tata Consultancy Services, India

Ashwini Patil
TCS Research, Tata Consultancy Services, India

Rinu Chacko
TCS Research, Tata Consultancy Services, India

Beena Rai
TCS Research, Tata Consultancy Services, India

ABSTRACT

The chemical industry is expanding its focus from process-centered products to product-centered products. Of these, consumer chemical products and other similar formulated products are especially ubiquitous. State of the art in the formulated product design relies heavily on experts and their expertise, leading to extended time to market and increased costs. The authors show that it is possible to construct a graph database of various details of products from textual sources, both offline and online. Similar to the "generate and test" approach, they propose that it is possible to generate feasible design variants of a given type of formulated product using the database so constructed. If they restrict the set of products that are applied to the skin, they propose to test the generated design variants using an in-silico model. Even though this chapter is an account of the work in progress, the authors believe the gains they can obtain from a readily accessible database and its integration with an in-silico model are substantial.

DOI: 10.4018/978-1-7998-0108-5.ch010

INTRODUCTION

The chemical industry is a diverse sector, with a vast range of processes and products. As per (Zhang et al., 2017), chemical engineering has been expanding its focus from process-centered products such as chlorine and ammonia to product-centered products such as mosquito repellent sprays. Gani and Ng (2015) classified chemical products into molecular products, formulated products, functional products, and devices. Such products include basic organic materials such as olefins, aromatics, biochemicals and plastics and basic inorganic materials such as engineered particles, inorganic chemicals, acids, gases produced from raw extracted materials and sustainable feedstocks. These basic chemicals, or ingredients, are used as the building blocks for the formulation of complex materials and substances such as speciality chemicals and consumer products known as formulated products.

Formulated products comprise a combination of raw materials engineered and designed to form powders, granules, tablets, creams, suspensions, foams, gels and emulsions, all displaying a set of targeted properties. Formulated products are ubiquitous in everyday applications such as lubricants, fuels, paints, inks, dyes, coatings, adhesives, detergents, cosmetics, personal care, household and professional care, medicines, foods, pesticides, construction materials, fuel additives and pharmaceutical products[1]. Individual ingredients used within a formulation may be incorporated to provide active functionality, enhanced delivery or as a protective or stabilizing agent (Chatterjee and Alvi, 2014).

Formulations of organic formulated products contain ingredients that undergo a step-by-step procedure such as heating, cooling, stirring, mixing, and so on to obtain specific target properties, both physical and chemical. A general understanding of the formulated products research area is that several types of data, methods and tools are necessary to tackle formulated product design problems to achieve an optimal design. Several approaches, conceptual models and frameworks have been proposed toward this goal (Bernardo and Saraiva, 2015; Conte et al., 2011; Gani and Ng, 2015; Hill, 2009; Lee et al., 2014; Martín and Martínez, 2013; Zhang et al., 2017; Zhang et al., 2018). Many approaches suggest using a database of relevant details for product design (Dionisio et al., 2018). The frameworks suggest using knowledge from experience, models or databases to choose a product form such as cream; then select types of ingredients such as solvent; generate candidates for each selected ingredient type and finally combine the ingredients (Zhang et al., 2017). Experts find similar formulations using standard file search and compilation. Lee et al. (2014) suggest a knowledge-based ingredient formulation system to support formulators in their attempt to select the most appropriate ingredients using past formulation cases. Lee et al. assert that without any knowledge support tools, chemical product development becomes iterative and time-consuming without a list of acceptable ingredients. In most of these approaches, the assumption is that either a relevant database/ knowledge-base is available or should be created manually.

We take the stance that it is possible to use the text of existing formulations to generate such a database. With a proper design of this database, it is also possible to make it highly searchable.

A vast amount of textual data is available occurring in sources such as textbooks, handbooks, journal articles, and specialized web sites. These texts are available in a form that makes at best a file search possible, but it is difficult to query the text flexibly. The information present in such texts contains a variety of domain-specific information such as the type of ingredients, mixtures of ingredients, functionalities and their compositions, and their physical attributes such as weights or weight fractions (Isaacs et al., 2016; Isaacs et al., 2018). In the search for a new formulation, an expert must refer to the already existing recipes to make rational judgments when choosing on the ingredients, their respective quantities and the procedure to follow to get a stable formulation that has the desired chemical function. This process