World Patent Information 35 (2013) 178–182

Contents lists available at SciVerse ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Markush structure searching by information professionals in the chemical industry – Our views and expectations

Peter Geyer

BASF SE, BASF Group Information Center, GV/WB – C006, 67056 Ludwigshafen, Germany

Keywords:
Leading opinion
Markush
Structure searching
Database developments
Database limitations
Searching problems
Indexing problems

A B S T R A C T

Markush structures are a special representation of the chemical compounds covered in patent documents. Due to their highly generic nature they are more difficult to index and search than specific chemical structures since they require special indexing and search features to make them searchable. Only a few databases exist that offer such features. The currently available indexing systems have been developed mainly in the 1960s to the 1980s, have not been majorly enhanced since then and their future seems somewhat uncertain. However, due to the fact that such Markush systems index unique information describing the chemistry protected by patent claims they are a necessary and indispensable information source for patent information professionals in the chemical industry in order to be able to support important business processes in chemical companies. Therefore we at BASF see a clear need to keep providing and also to keep developing such systems in the future.

1. Introduction

A certain type of representation of chemical structures used in patent documents, the so-called Markush structures, is of particular importance for the chemical and pharmaceutical industry. Markush structures are a generic type of description of chemicals used to summarize a potentially very large set of closely related chemicals in a single condensed representation. They are used in descriptions and especially in claims of patent documents and are a key feature of these documents describing the chemistry protected by them. Markush structures were named after Dr. Eugene Markush who was the first inventor to successfully use such a type of representation in patent claims at the USPTO in 1924 [1]. Markush structures consist of a "core" chemical structure and a list of possible substituents (commonly called R-groups) attached to it. They use four types of possible substituent options: Substituent variation (allowing different substituents at a position), position variation (allowing different attachment points for a substituent), frequency variation (allowing substituents to occur multiple times) and homology variation (using generic expressions covering many specific substituents like "alkyl") [2]. Markush structures used in patents tend to get more and more complex and a single Markush description can today cover several pages of a patent document. In order to make such structures available for retrieval in patent searches different types of indexing systems were developed in the 1960s to the 1980s (for a review of these systems and their development see Refs. [3–5]): The older fragmentation code systems including the Derwent (now Thomson-Reuters) CPI fragmentation code [6], the IFI Claims® code [7] and the GREMAS code [8] use text codes to describe different features of a Markush structure. Due to the merely text-based nature of these codes the correct connection of the fragments in the structure typically cannot fully be reproduced in these systems, making searches less precise. The newer and more specific topological search systems include Marpat® from Chemical Abstracts Service [9,10] and the Merged Markush Service (MMS) jointly developed by Derwent, the French INPI and Questel [11]. These systems are able to describe Markush structures in a graphical form similar (but not identical) to the description in the original patent document including the full connectivity of the single atoms/fragments allowing more precise searches. However, all these indexing systems are many years old, have not much been developed further since their introduction and are quite complicated to use.

2. Why Markush search systems are necessary

Apart from these Markush search systems other well established systems exist for searching the chemistry described in patent documents. These include e.g. databases like CAS Registry℠ from Chemical Abstracts Service [12], the Derwent Chemistry Resource in the Derwent World Patents Index® from Thomson-Reuters [13] or Beilstein/Reaxys® from Elsevier [14]. Such databases contain specific chemicals described in literature and/or patent documents.

E-mail address: peter.geyer@basf.com.

0172-2190/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wpi.2013.05.002

For patents they are mainly focused on chemicals described in the examples and/or claims sections. For many years these databases have been commonly used by chemical information professionals. In recent years also further systems and technologies have become available indexing chemistry contained in patent documents that previously had not been available for retrieval. For example, Chemical Abstracts Service has recently started indexing so-called "prophetic" chemicals described in patent documents in Registry[SM], i.e. specific chemicals disclosed in the documents but without the description of an actual synthesis [15]. Apart from that several providers offer systems containing specific chemicals from patent documents derived by automated means from the documents, for instance Surechem [16] or IBM [17], or the respective technologies for obtaining these, i.e. text and/or image mining technologies (for a review see Ref. [18]).

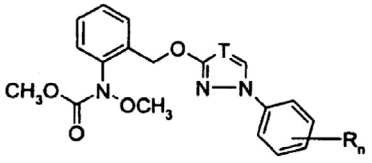
With all these possibilities for searching patent chemicals and in light of the difficulties of using the Markush search systems why are these Markush systems still so important for patent information professionals in the chemical industry? The main reason is that these systems are necessary for comprehensively searching the chemistry claimed in patent documents since Markush structures are typically used in chemical patent claims. Usually specific chemicals belonging to the chemical space described by the Markush structures of the claims are described as examples (including prophetics) in the description of the patents. However, these specific chemicals typically cover only a part of the chemical space covered by the Markush structures in the respective claims. In addition, the mentioned systems containing specific chemicals of patent documents are to date not even comprehensively covering all the specific chemicals described in these documents. For the traditional indexing databases chemical descriptions in the documents have to match certain indexing guidelines in order to get indexed (this is even true for the mentioned indexing of the prophetic chemicals). Also the automated methods are not yet capable of comprehensively extracting all specific chemicals from the documents. This is especially true in the case of the specific chemicals disclosed in tables requiring a combination of image mining (for the graphical core structure and graphical substituents) and text mining (for the text representation of substituents in the table). Finally, if a patent is covering a chemical of interest only by its Markush claims and not through a disclosure of specific chemicals, Markush search systems are the only way to find such a patent in a chemical structure search. Therefore, especially for freedom-to-operate questions (i.e. would an intended action infringe patents of 3rd parties?) these systems remain a core information source for chemical information professionals. The situation explained above can be illustrated by the example described in the following.

Let's imagine a chemical company would like to market the mixture of the insecticide Fipronil with the fungicide Pyraclostrobin. This mixture is claimed by the European patent EP 1696728 B1 [19]. The document's Markush structures in the claims are even not very broad, at least compared to some other chemical patent claims. The claims cover only a limited set of combinations of the two mentioned compounds and other closely related chemicals (see Fig. 1). The company planning to market the mixture therefore

Claims

1. A mixture for crop protection, comprising as active components

a) carbamate derivatives of the formula I

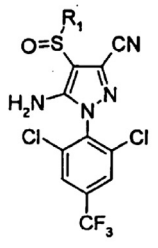


I

in which the substituents and the index have the following meaning:

TisCHorN
n is 0, 1 or 2
R is halogen, C_1-C_4-alkyl or C_1-C_4-haloalkyl, it being possible for the radicals R to be different when n is 2,

and
b) at least one compound of the formulae II



II

in which R_1 is C_1-C_4-alkyl or C_1-C_4-haloalkyl;
in a synergistic effective amount.

Fig. 1. Claim 1 of EP1696728B1.