



HAL
open science

On Pattern Setups and Pattern Multistructures

Aimene Belfodil, Sergei O. Kuznetsov, Mehdi Kaytoue

► **To cite this version:**

Aimene Belfodil, Sergei O. Kuznetsov, Mehdi Kaytoue. On Pattern Setups and Pattern Multistructures. 2019. hal-02147757

HAL Id: hal-02147757

<https://hal.archives-ouvertes.fr/hal-02147757>

Preprint submitted on 5 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Pattern Setups and Pattern Multistructures

Aimene Belfodil^{a, c}, Sergei O. Kuznetsov^b, Mehdi Kaytoue^{a, d}

^aUniv Lyon, INSA Lyon, CNRS, LIRIS UMR 5205, F-69621, LYON, France; ^bNational Research University Higher School of Economics, Moscow, Russia; ^cMobile Devices Ingénierie, 100 Avenue Stalingrad, 94800, Villejuif, France; ^dInfologic, 99 avenue de Lyon, 26500 Bourg-Lès-Valence, France;

ARTICLE HISTORY

Compiled June 5, 2019

ABSTRACT

Modern order and lattice theory provides convenient mathematical tools for pattern mining, in particular for condensed irredundant representations of pattern spaces and their efficient generation. *Formal Concept Analysis (FCA)* offers a generic framework, called *pattern structures*, to formalize many types of patterns, such as itemsets, intervals, graph and sequence sets. Moreover, FCA provides generic algorithms to generate irredundantly all closed patterns, the only condition being that the pattern space is a meet-semilattice. This does not always hold, e.g., for sequential and graph patterns. Here, we discuss *pattern setups* consisting of descriptions making just a partial order. Such a framework can be too broad, causing several problems, so we propose a new model, dubbed *pattern multistructure*, lying between pattern setups and pattern structures, which relies on *multilattices*. Finally, we consider some techniques, namely *completions*, transforming pattern setups to pattern structures using sets/antichains of patterns.

1. Introduction

Modern order and lattice theory provide convenient mathematical tools for pattern mining, in particular for condensed irredundant representations of pattern spaces and their efficient generation. Different formal tools has been proposed in the literature to model pattern spaces. Formal Concept Analysis (FCA - Ganter and Wille (1999)) has been proposed by Wille (1982) as a well-founded mathematical tool to models hierarchies of concepts related to some formal context (i.e. binary datasets). While basic FCA provides a natural way to analyze binary datasets, datasets with more complex attributes (e.g. numerical or nominal ones) need to be transformed to such before any manipulation. This kind of transformation has been proposed by Ganter and Wille (1989) under the term of *conceptual scaling* (i.e. *binarizing*). Yet, even if *conceptual scaling* is a quite general tool, *binarizing* a dataset with regard to some pattern search spaces is not always obvious (i.e. Baixeries, Kaytoue, and Napoli (2012), Belfodil et al. (2017)). In response to that, some other more natural tools to formalize complex pattern spaces has been proposed. One could cite Logical Concept Analysis (LCA) proposed by Ferré and Ridoux (2000) and Pattern Structures proposed by Ganter and Kuznetsov (2001). Pattern Structures allow for instance to model in a quite natural

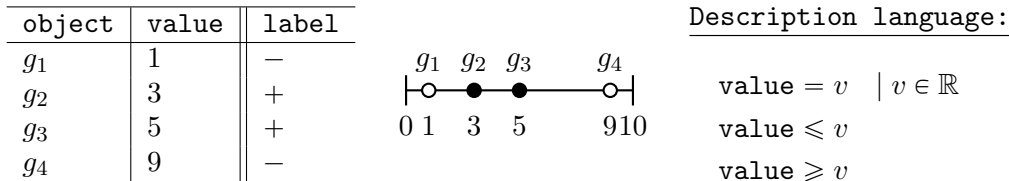


Figure 1. Dataset with numerical attribute (left), its representation in \mathbb{R} with black dots representing positive objects (center) and the description language (right).

way many pattern search spaces. Indeed, itemsets, intervals (Kaytoue, Kuznetsov, and Napoli 2011), convex polygon (Belfodil et al. 2017), partition (Baixerries, Kaytoue, and Napoli 2014) pattern spaces among others (Ganter and Kuznetsov 2001; Kuznetsov 2009) can be modeled within the pattern structure framework. However, since pattern structures rely on meet-semilattices (i.e. conjunction of two patterns belongs to the pattern search spaces), some pattern spaces that are only partially ordered sets (posets) cannot be “directly” defined using such a framework.

Consider for instance the example dataset depicted in Fig. 1 containing 4 objects described by attribute “**value**” and labeled positive or negative. We are interested by the task of finding “good” rules $d \rightarrow +$ in this dataset with d a description given by attribute **value**. Rather than considering the usual meet-semilattice of intervals as the one proposed by Kaytoue, Kuznetsov, and Napoli (2011); descriptions d are restrained to open intervals of the form $(v]$ and $[v)$ or singleton $\{v\} \subseteq \mathbb{R}$ (see Fig. 1 - right). Patterns (descriptions) form together a poset (\mathcal{D}, \supseteq) where \supseteq is the *subsumption order* (i.e. if d_1 subsumes d_2 then if pattern d_2 holds for an object g then pattern d_1 hold to). However, (\mathcal{D}, \supseteq) does not form a meet-semilattice. For instance, the set $\{\{3\}, \{5\}\}$ does not have a meet, since lower bounds of $\{\{3\}, \{5\}\}$ have two maximal elements w.r.t. \supseteq (i.e. $[3)$ and $(5]$). Hence, the description space does not induce a pattern structure (Ganter and Kuznetsov 2001). It does form actually a *pattern setup* (Lumpe and Schmidt (2015)) which rely on a simple poset with no additional properties.

Description spaces like the one depicted in Fig. 1 are numerous. For instance, sequence of itemset patterns (Agrawal and Srikant (1995)) ordered by “*is subsequence of*” do not form a meet-semilattice (Zaki 2001; Codocedo et al. 2017). Sequential meet-semilattice in FCA (Buzmakov et al. 2016; Codocedo et al. 2017) refers usually to *set (i.e. conjunction) of sequences* rather than to the poset of *sequences* itself. Same holds for the graph meet-semilattice from Kuznetsov (1999). In general, the base pattern setup is transformed to a pattern structure using sets (i.e. conjunctions) of descriptions thus providing a richer pattern language. Such transformations are naturally called *completions*, which refers to well-known completions in ordered sets and lattice theory, like Dedekind-McNeille, Alexandroff or Antichain completions (Davey and Priestley 2002; Boldi and Vigna 2016). For instance, in Fig. 1, set of patterns $\{\mathbf{value} \geq 3, \mathbf{value} \leq 5\}$ which is equivalent to $3 \leq \mathbf{value} \leq 5$ belongs to the antichain completion of (\mathcal{D}, \supseteq) but and not to the base description language (\mathcal{D}, \supseteq) .

Understanding properties of pattern setups independently from their completions is fundamental for answering many practical questions. For instance, consider the question “*What are the best descriptions covering all positive instances?*”. If *better* stands for *more relevant than* as in relevance theory (Garriga, Kralj, and Lavrac 2008), the answer will be the two best incomparable rules $\mathbf{value} \geq 3 \rightarrow +$ and $\mathbf{value} \leq 5 \rightarrow +$ rather than only one in the completion $3 \leq \mathbf{value} \leq 5 \rightarrow +$. More generally, we look for *maximal common patterns* (i.e. *support-closed patterns* as called by Boley et al. (2010)) of the positive instances which could be multiple.

Outlines and Contributions. In this paper we present the following results:

- (1) In Section 3, we start by studying pattern setups that was proposed by Lumpe and Schmidt (2015) as a tool to models pattern search spaces relying only on posets providing hence a better-understanding of this wider-class of framework. Afterwards, we revisit briefly the notion of pattern structures in Section 4.
- (2) In Section 5, we point-out the major problem related to pattern setups. In fact, since pattern setups rely only on posets without no additional properties, they are very permissive. Simply put, given some set of objects, they could share some common descriptions in the pattern space but none of them is maximal w.r.t. the subsumption order. We show that this problem is directly linked to the fact that the considered poset is not a multilattice.
- (3) In Section 6, we present multilattices and we show that all (doubly) chain-complete posets are complete multilattices but that the converse does not hold.
- (4) In Section 7, we propose the framework of pattern multistructures. In a nutshell, analogously to pattern structures, pattern multistructures are based on multilattices. Such a structure provides the fact that covering descriptions of a set of objects are deducible from the maximal common descriptions using the subsumption order. This does not necessarily hold in an arbitrary pattern setup.
- (5) Next, in Section 8 we revisit completion (i.e. transformation) of pattern setups to pattern structures and we show that the usual completion using antichain of patterns, namely *antichain completion*, induce a pattern structure if and only if the pattern setup is a pattern multistructure.
- (6) Finally, we wrap-up in section 9 and we discuss some open problems related particularly to the enumeration of extents (i.e. subset of objects that are separable in the description language) in an arbitrary pattern setup.

Please note that this paper is a thorough extension of a first paper published in CLA'18 (Belfodil, Kuznetsov, and Kaytoue 2018). The main difference with CLA'18 paper is more mathematical results and much more detailed explanations of the framework of pattern setups, pattern multistructures and their completions. Note that in this first paper, pattern multistructures were called pattern hyper-structures and were linked with the notion of hyper-lattices briefly investigated by Zaki (2001). Stefan Schmidt attracted our attention to the work of Benado (1955) on multilattices that interestingly was directly linked to pattern multistructure making this newly proposed framework more mathematically founded.

2. On Partially-Ordered Sets

To make this work self-contained, in this section we recall basic definitions and results from order theory and introduce our notations that are largely inspired by Ganter and Wille (1999); Davey and Priestley (2002); Roman (2008). In this paper, we will be using the following base notations:

- For any set P , the set $\wp(P)$ denotes the powerset of P .
- For any mapping $f : E \rightarrow F$ and any subset $A \subseteq E$, the set $f[A]$ denotes the image of A w.r.t. f , that is:

$$f[A] = \{f(a) \mid a \in A\}$$

- The set of all mappings $f : E \rightarrow F$ is denoted by F^E .

2.1. Basic Definitions

Definition 2.1. A **partial order** on a set P is a binary relation \leq on P that is **reflexive** ($(\forall x \in P) x \leq x$), **transitive** ($(\forall x, y, z \in P)$ if $x \leq y$ and $y \leq z$ then $x \leq z$.) and **antisymmetric** ($(\forall x, y, z \in P)$ if $x \leq y$ and $y \leq x$ then $x = y$). The pair (P, \leq) is called a **partially ordered set** or a **poset** for short. Two elements x and y from P are said to be **comparable** if $x \leq y$ or $y \leq x$; otherwise, they are said to be **incomparable**. A subset $S \subseteq P$ is said to be a **chain** (resp. an **antichain**) if all elements of S are pairwise comparable (resp. incomparable). The set of all chains (resp. antichains) of P is denoted by $\mathcal{C}(P)$ (resp. $\mathcal{A}(P)$).

Note 1. (Finite) posets are generally presented using the so called **Hasse Diagram**. Where elements of the poset are represented on the plane from the smallest ones to the largest ones. Only direct neighbors (i.e. two elements are said to be direct neighbors if there is no elements strictly lying between them in the poset) are linked by a segment of line. The poset can indeed be deduced from this diagram by adding reflexivity and transitivity. Fig. 2 (1) depicts the Hasse Diagram of the powerset of a set $E = \{a, b, c\}$ ordered by set inclusion (i.e. poset $(\wp(E), \subseteq)$).

Through some misuse of the notion, we will present in this paper some infinite posets using the Hasse Diagram. Fig. 2 (2) presents a poset (P, \leq) where:

- $P = \{\perp, \top, b_0, b_1\} \cup \{a_i \mid i \in \mathbb{N}\}$,
- $\perp \leq b_0 \leq b_1 \leq \top$, and
- $(\forall i \in \mathbb{N}) \perp \leq a_i \leq a_{i+1} \leq \top$.

In what follows, (P, \leq) denotes a poset and $S \subseteq P$ denotes an arbitrary subset.

Definition 2.2. The **principal ideal** (resp. **principal filter**) of $p \in P$, denoted by $\downarrow p$ (resp. $\uparrow p$) is the set of all elements below (resp. above) it:

$$\downarrow p = \{x \in P \mid x \leq p\} \quad \uparrow p = \{x \in P \mid p \leq x\}$$

Definition 2.3. A set $S \subseteq P$ is said to be a **lower ideal** or a **downset** if:

$$(\forall x \in P) ((\exists s \in S) x \leq s \Rightarrow x \in S)$$

The notion of **upper ideal** (**upset**) is defined dually. The set of all lower (resp. upper) ideals of (P, \leq) is denoted $\mathcal{O}(P)$ (resp. $\mathcal{U}(P)$).

Definition 2.4. The **down closure** (resp. **up closure**), denoted by $\downarrow S$ (resp. $\uparrow S$) associates to a subset S the smallest lower-ideal (resp. upper-ideal) enclosing it:

$$\downarrow S = \{x \in P \mid (\exists s \in S) x \leq s\} = \bigcup_{s \in S} \downarrow s \quad \uparrow S = \{x \in P \mid (\exists s \in S) s \leq x\} = \bigcup_{s \in S} \uparrow s$$

Definition 2.5. An element $p \in P$ is said to be a **lower bound** (resp. **upper bound**) of S if it is below (resp. above) all elements of S . The set of lower (resp. upper) bounds of S in P , denoted by S^ℓ (resp. S^u), is the *lower ideal* (resp. *upper ideal*) given by:

$$S^\ell = \{x \in P \mid (\forall s \in S) x \leq s\} = \bigcap_{s \in S} \downarrow s \quad S^u = \{x \in P \mid (\forall s \in S) s \leq x\} = \bigcap_{s \in S} \uparrow s$$

Note 2. Note that $\uparrow \emptyset = \downarrow \emptyset = \emptyset$ and $\emptyset^\ell = \emptyset^u = P$.

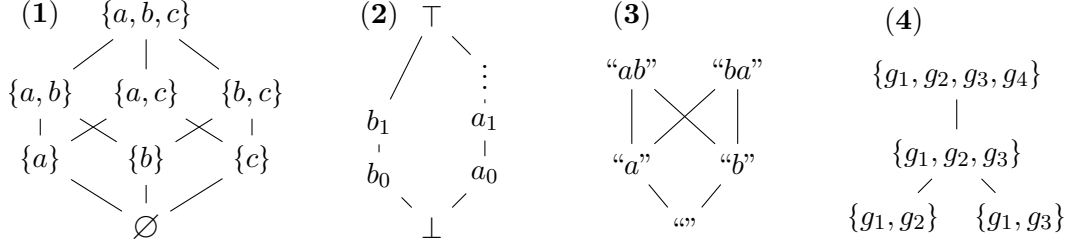


Figure 2. From left to right: (1) Poset $(\wp(\{a, b, c\}), \subseteq)$. (2) A poset (P, \leq) with $P = \{\perp, \top, b_0, b_1\} \cup \{a_i \mid i \in \mathbb{N}\}$, $\perp \leq b_0 \leq b_1 \leq \top$ and $(\forall i \in \mathbb{N}) \perp \leq a_i \leq a_{i+1} \leq \top$. (3) Poset of common substrings of “ab” and “ba”. (4) A subposet of $(\wp(\{g_1, g_2, g_3, g_4\}), \subseteq)$.

2.2. Minimal, Minimum, Infimum, ...

Informally speaking, not all elements in a subset S have the same “position”. Some elements x are in the “interior” of S , that is $\exists y, z \in S$ such that $y < x < z$; while some others are on its “border”, that is there is no element in S strictly below/above them. These border elements are formally defined in Definition 2.6.

Definition 2.6. An element $s \in S$ is said to be a **minimal** (resp. **maximal**) element in S if all its strict lower (resp. upper) bounds are outside S . The set of *minimal* (resp. *maximal*) elements of S is denoted by $\min(S)$ (resp. $\max(S)$) and given by:

$$\min(S) = \{x \in S \mid \downarrow x \cap S = \{x\}\} \quad \max(S) = \{x \in S \mid \uparrow x \cap S = \{x\}\}$$

Note that $\min(S)$ and $\max(s)$ are *antichains*. Moreover, $\min(\emptyset) = \max(\emptyset) = \emptyset$ by definition. An intuitive and important relationship between the minimal elements of a subset and the minimal elements in its up-closure is presented lemma 2.7.

Lemma 2.7. We have $\min(\uparrow S) = \min(S)$ and $\max(\downarrow S) = \max(S)$

Proof. We prove by double inclusion the property $\min(\uparrow S) = \min(S)$:

(\subseteq) Let $x \in \min(\uparrow S)$ and suppose that $x \notin S$. Since $x \in \min(\uparrow S)$ we have $x \in \uparrow S$, that is $\exists y \in S$ s.t. $y \leq x$ but $y \neq x$ since $x \notin S$. Thus, $\exists y \in \uparrow S$ s.t. $y \leq x$ but $y \neq x$. Thus $y \in \downarrow x \cap \uparrow S$ with $y \neq x$ which contradicts the fact that $x \in \min(\uparrow S)$ (i.e. $\downarrow x \cap \uparrow S = \{x\}$). We conclude that $x \in S$. Suppose now that $x \notin \min(S)$, that $\exists y \in S$ s.t. $y \leq x$ and $y \neq x$. Hence, $y \in \uparrow S \cap \downarrow x$ which contradicts the fact that $x \in \min(\uparrow S)$. Thus $x \in \min(S)$. We conclude that $\min(\uparrow S) \subseteq \min(S)$.

(\supseteq) Let $x \in \min(S)$, thus $x \in \uparrow S$. Suppose that $x \notin \min(\uparrow S)$ that is $\exists y \in \uparrow S$ such that $y \leq x$ and $y \neq x$. Thus $\exists z \in S$ such that $z \leq y \leq x$ with $z \neq x$. Hence, $z \in \downarrow x \cap S$ with $z \neq x$ which contradicts the fact that $x \in \min(S)$. We conclude that $x \in \min(\uparrow S)$ or more generally $\min(S) \subseteq \min(\uparrow S)$.

One can follow the same steps to show $\max(\downarrow S) = \max(S)$. \square

Definition 2.8. An element $m \in S$ is said to be **minimum** or a **smallest element** of S if it is a lower bound of S . Formally: $m \in S$ and $(\forall s \in S) m \leq s$. The notion of **maximum** or **greatest element** of S is defined dually.

Note 3. One should notice that not all subsets have a minimum or a maximum (see Example 2.9). Moreover, the minimum and the maximum are unique if they exist.

Example 2.9. Consider again Fig. 2 (1) and the subset $S = \{\{a\}, \{a, b\}, \{a, c\}\}$. It is clear that $\min(S) = \{a\}$ and $\max(S) = \{\{a, b\}, \{a, c\}\}$. Obviously, subset S has a minimum which is $\{a\}$. However, it does not have a maximum since no element in S is above all the other elements.

Note 4. If a poset has a maximum (resp. minimum) element, we say that the poset is **upper-bounded** (resp. **lower-bounded**) and this element is called the **top** (resp. **bottom**) and is denoted \top (resp. \perp). A poset is said to be **bounded** if it is both lower-bounded and upper-bounded.

It is very important to distinguish between the minimal elements and the minimum. In fact, “if a subset has a minimum then it has a unique minimal element”. However, the converse of this statement is not true, that is even if a subset have a single minimal element, it can have no minimum. Indeed, the intuition that every element in a subset S is at least above one minimal element in $\min(S)$ does not always hold in infinite posets. Such a property is presented in Definition 2.10.

Definition 2.10 (Definition 2.4 in Martínez et al. (2005)). We say that:

- S is **minimum-handle** if S has a minimum (i.e. $(\exists m \in S) S \subseteq \uparrow m$).
- S is **maximum-handle** if S has a maximum (i.e. $(\exists m \in S) S \subseteq \downarrow m$).
- S is **minimal-handle** if $S \subseteq \uparrow \min(S)$
- S is **maximal-handle** if $S \subseteq \downarrow \max(S)$.

Example 2.11. Consider the poset (P, \leq) depicted in Fig. 2 and let $S = P \setminus \{\top\} = \{\perp, b_0, b_1\} \cup \{a_i \mid i \in \mathbb{N}\}$. It is clear that $\max(S) = \{b_1\}$. Since $\downarrow \{b_1\} = \{\perp, b_0, b_1\}$, subset S is not a maximal-handle. Hence, even if S has a single maximal element, it has no maximum. On the other hand, $\min(S) = \perp$. Since $S \subseteq \uparrow \perp = P$, we can say that S is a minimal-handle. Moreover, since the minimal element is unique than S is minimum-handle which minimum is \perp .

Note 5. If S is minimal-handle, then S has a minimum if and only if S has a unique minimal element. It is clear that all subsets of a **finite posets** (i.e. posets with a finite set) are maximal-handle and minimal-handle. Posets where all subsets are minimum-handle are said to be **well-founded** or equivalently have the **minimal condition** or the **descending chain condition (DCC)**. Dually, posets where all subsets are maximal-handle are said to be **dually well-founded** or equivalently have the **maximal condition** or the **ascending chain condition (ACC)**. A poset having at the same time the **ACC** and **DCC** is said to be **chain-finite** since it has no infinite chain (but, still, could have infinite antichains). A poset is in fact finite if and only if it is **chain-finite** and **antichain-finite**.

Definition 2.12. The largest lower bound of S (i.e. the maximum of S^ℓ) if it exists is called the **infimum** or the **meet** of S and is denoted $\inf(S)$ or $\bigwedge S$. The **join** or the **supremum** of S is given by the minimum of S^u and is denoted $\sup(S)$ or $\bigvee S$.

Having a minimum is a very strong property, in fact if a subset has a minimum then the minimum is also the infimum. However, the converse is not true. Indeed a subset can have an infimum without having a minimum. For instance, consider Fig. 2 (2), it is clear that the chain $S = \{a_i \mid i \in \mathbb{N}\}$ does not have a maximum. However, S still have a supremum which is \top . Indeed, $S^u = \{\top\}$ which minimum is \top .

One interesting property is stated in the following Lemma.

Lemma 2.13. Let (P, \leq) be a poset, $S \subseteq P$, we have:

- For any $A \subseteq S^\ell$. If A has a join $\bigvee A \in P$ then $\bigvee A \in S^\ell$.
- For any $A \subseteq S^u$. If A has a meet $\bigwedge A \in P$ then $\bigwedge A \in S^u$.

Proof. Let $A \subseteq S^\ell$, we have by definition: $(\forall s \in S \forall a \in A) a \leq s$, that is $S \subseteq A^u$. Since $\bigvee A$ is the least upper bound of A and all elements of S are upper bounds of A then: $(\forall s \in S) \bigvee A \leq s$. We conclude that $\bigvee A \in S^\ell$. Same steps can be followed to show the second part of the Lemma. \square

Note 6. One should note that, in case of existence, we have:

$$\bigwedge S = \bigvee S^\ell \quad \bigvee S = \bigwedge S^u$$

Moreover, since for a poset (P, \leq) we have $\emptyset^\ell = \emptyset^u = P$, then the empty set has a meet (resp. join) if and only if the poset is upper-bounded (resp. lower bounded) and we have $\bigwedge \emptyset = \bigvee P = \top$ and $\bigvee \emptyset = \bigwedge P = \perp$.

2.3. Lattices

Definition 2.14. A poset (P, \leq) is said to be:

- A **meet-semilattice** if for all nonempty finite subsets $S \subseteq P$, S has its meet.
- A **join-semilattice** if for all nonempty finite subsets $S \subseteq P$, S has its join.
- A **lattice** if it is both meet-semilattice and join-semilattice.
- A **complete lattice** if all its subsets including \emptyset has their meet and join.

Note 7. For any set E , the poset $(\wp(E), \subseteq)$ is a complete lattice where the meet is set intersection \cap and the join is set union \cup . Such a poset is called a **powerset lattice**.

Example 2.15. Consider posets depicted in Posets Fig. 2. Poset **(1)** is a powerset lattice on $\{a, b, c\}$. Hence, it is a complete lattice. Poset **(2)** is also a complete lattice where the join of any infinite subset is \top . Poset **(3)** of common substrings of “**ab**” and “**ba**” ordered by *is substring of* is neither a meet-semilattice nor a join-semilattice. Indeed, $\{“ab”, “ba”\}$ has not an infimum since $\{“ab”, “ba”\}^\ell = \{“a”, “b”, “”\}$ has two maximal elements and thus no maximum. Dually, subset $\{“a”, “b”\}$ has not a join since $\{“a”, “b”\}^u = \{“ab”, “ba”\}$ has two minimal elements and thus no minimum.

Note 8. Meet-semilattices have a weaker, yet equivalent, definition characterizing them. In fact, to check if a poset is a lattice, one should only check if:

$$(\forall x, y \in P) \{x, y\} \text{ has a meet} \quad \mathbf{and} \quad (\forall x, y \in P) \{x, y\} \text{ has a join}$$

That is if all pair of elements have their meets (resp. join) then all nonempty finite subsets of the poset have their meet (resp. join). Another important remark that is related to Note 6 is the fact that all complete semilattices are complete lattices.

One should notice that since in complete lattices, the empty set have also their meet and joins then all complete lattices are bounded. It is also important to note that all finite lattices are complete lattice. However, not all finite meet-semilattices are lattices since they may lack of a top element (i.e. the meet of the empty set is not guaranteed to exist). In fact, a finite meet-semilattice (resp. join-semilattice) is a lattice if and only if it is upper-bounded (resp. lower-bounded).

2.4. Morphisms on Posets

We will often use morphismes (i.e. mappings) between posets in this paper. Definition 2.16 below formulate some properties of morphismes between two posets.

Definition 2.16. Let (P, \leq) and (Q, \leq) be the two posets. A mapping $f : P \rightarrow Q$ is:

- **order-preserving** or **monotone**: $(\forall x, y \in P) x \leq y \Rightarrow f(x) \leq f(y)$.
- **order-reversing**: $(\forall x, y \in P) x \leq y \Rightarrow f(y) \leq f(x)$
- **an order-embedding**: $(\forall x, y \in P) x \leq y \Leftrightarrow f(x) \leq f(y)$

If an order-embedding exist from (P, \leq) and (Q, \leq) then poset (Q, \leq) is said to be a **completion** of (P, \leq) or **embeds** (P, \leq) . If this order-embedding is surjective (i.e. an **order-isomorphism**) we say that (P, \leq) and (Q, \leq) are **order-isomorphic**.

Definition 2.17. A **closure operator** on (P, \leq) is a mapping $\phi : P \rightarrow P$ that is:

- **monotone**: $(\forall x, y \in P) x \leq y \Rightarrow \phi(x) \leq \phi(y)$,
- **extensive**: $(\forall x \in P) x \leq \phi(x)$, and
- **idempotent**: $(\forall x \in P) \phi(\phi(x)) = \phi(x)$

Note 9. The **fixpoints** of a given mapping $f : P \rightarrow P$ is the set of elements s.t. $\{p \in P \mid f(p) = p\}$. For idempotent operator, the set of fixpoints is $f[P] = \{f(p) \mid p \in P\}$.

Lemma 2.18. Let (P, \leq) be a poset, we have \uparrow and \downarrow are closures on $(\wp(P), \subseteq)$.

Proof. Let us show that $\uparrow: \wp(P) \rightarrow \wp(P)$ is a closure operator on $(\wp(P), \subseteq)$. It is clear that $S \subseteq \uparrow S$ (i.e. \uparrow is **extensive**) by definition. For $S \subseteq T$ in $\wp(P)$, we have if $x \in \uparrow S$, then $\exists y \in S \subseteq T$ such that $y \leq x$. That is, $x \in \uparrow T$. Thus, $(\forall S, T \in \wp(P)) \uparrow S \subseteq T \implies \uparrow S \subseteq \uparrow T$ (i.e. \uparrow is **order-preserving**). Let us show that \uparrow is **idempotent**. It is clear that $\uparrow S \subseteq \uparrow \uparrow S$ since \uparrow is extensive. It remains to show that $\uparrow \uparrow S \subseteq \uparrow S$. Let $x \in \uparrow \uparrow S$, that is $\exists y \in \uparrow S$ such that $y \leq x$. That is $\exists z \in S$ such that $z \leq y \leq x$. We conclude that $x \in \uparrow S$. One can follow the same steps to show that $\downarrow: \wp(P) \rightarrow \wp(P)$ is a closure operator on $(\wp(P), \subseteq)$. \square

Note 10. For any poset (P, \leq) and since \uparrow and \downarrow are closure operators, we have:

- For all $S \in \mathcal{U}(P)$, if S is minimal-handle then $S = \uparrow \min(S)$.
- For all $S \in \mathcal{O}(P)$, if S is maximal-handle then $S = \downarrow \max(S)$.

Before finishing the section, we draw the reader attention to the tight link existing between closure operators and closure systems on a complete lattice defined below.

Definition 2.19. Let (P, \leq) be a complete lattice whose meet is \bigwedge . Given a subset $S \subseteq P$, (S, \leq) is said to be a **closure system** or a **meet-structure** on (P, \leq) iff:

$$(\forall A \subseteq S) \bigwedge A \in S$$

Clearly, (S, \leq) forms a complete lattice which infimum is \bigwedge .

In fact (see Theorem 1 in Ganter and Wille (1999)), if (P, \leq) is a complete lattice and ϕ is a closure operator on (P, \leq) then the poset $(\phi[P], \leq)$ of fixpoints of ϕ is a *closure system*. Conversely, if (S, \leq) is a closure system on a complete lattice (P, \leq) then the mapping $\phi_S : P \rightarrow S, p \mapsto \bigwedge \{s \in S \mid p \leq s\}$ that takes each element $p \in P$ to the smallest element (fixpoint) in S enclosing it is a closure operator with $\phi_S[P] = S$.

3. Pattern Setups

Formal Concept Analysis (FCA) were introduced in Wille (1982) as a mathematical framework to analyze and manipulate concepts in databases. *FCA* starts by a *formal context* $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ where \mathcal{G} is a set of objects (i.e. *Gegenstände*), \mathcal{M} is a set of attributes (i.e. *Merkmale*) and \mathcal{I} is a binary relation on $\mathcal{G} \times \mathcal{M}$ (i.e. *Incidence relation*). For $(g, m) \in \mathcal{G} \times \mathcal{M}$, $g \mathcal{I} m$ holds iff g has attribute m . Fig. 3 presents an example of a formal context. The basic theorem behinds *FCA* rely on the observation that any formal context can be transformed to a complete lattice called *concept lattice*. We invite the reader to read Ganter and Wille (1999) book for more details.

While (basic) *FCA* gives a tool to analyze datasets in a form of formal context, datasets with more complex attributes (eg. numerical or nominal attributes) needs to be transformed to such a form before any manipulation. Such a transformation is called *conceptual scaling* (i.e. *binarizing*) (Ganter and Wille (1989)). Yet, even if *conceptual scaling* is a quite general tool, binarizing a dataset with regard to the patterns we want to look for is not always straightforward (Baixeries, Kaytoue, and Napoli (2012)).

In response to that, a more natural way to handle complex datasets was introduced in Ganter and Kuznetsov (2001) under the name of *pattern structures*. Objects in a pattern structure have descriptions (e.g. the equivalent notion to set of attributes in \mathcal{M} in a formal context) with a meet-semilattice operation on them (e.g. equivalent to set intersection in $(\wp(\mathcal{M}), \subseteq)$ in a formal context). This framework proved its usefulness in many data analysis tasks (see Kuznetsov (2009)). However, pattern structures demands that the *description space* to be a (upper-bounded) meet-semilattice which is not the case for all description spaces such as sequence of itemsets patterns (Codocedo et al. (2017)). *Pattern setups* were introduced in Lumpe and Schmidt (2015) to generalize pattern structures by demanding only a partial order on descriptions. We details in this sections the different notions related to pattern setups.

Definition 3.1. A **description space**; called also **description language**, **pattern space** or **pattern language**; is any *poset* $\underline{\mathcal{D}} := (\mathcal{D}, \sqsubseteq)$. Elements of \mathcal{D} are called **descriptions** or **patterns**. For any $c, d \in \mathcal{D}$, $c \sqsubseteq d$ should be read as “ c is *less restrictive than* d ” or “ c **subsumes** d ”.

Definition 3.2. A **pattern setup** is a triple $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ where \mathcal{G} is a *set* (of objects), $\underline{\mathcal{D}}$ is a description space and $\delta : \mathcal{G} \rightarrow \mathcal{D}$ defines a mapping that takes each object $g \in \mathcal{G}$ to its description $\delta(g) \in \mathcal{D}$. Let $g \in \mathcal{G}$ and $d \in \mathcal{D}$ be an object and a description, respectively. We say that object g **realizes** description d or description d **hold for** or **cover** object g iff $d \sqsubseteq \delta(g)$.

Example 3.3. Consider the pattern setup $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ in Fig. 4. We have $\mathcal{G} = \{g_i\}_{1 \leq i \leq 4}$. The description space is the set of *nonempty words* on the alphabet $\{a, b, c\}$ (i.e. $\{a, b, c\}^+$) ordered by the relationship “*is substring of*” \sqsubseteq . The mapping δ asso-

\mathcal{G}	a	b	c
g_1	×	×	×
g_2	×		
g_3	×		
g_4		×	×

Figure 3. Formal Context $(\mathcal{G}, \mathcal{M}, \mathcal{I})$ with $\mathcal{G} = \{g_i\}_{1 \leq i \leq 4}$ and $\mathcal{M} = \{a, b, c\}$.

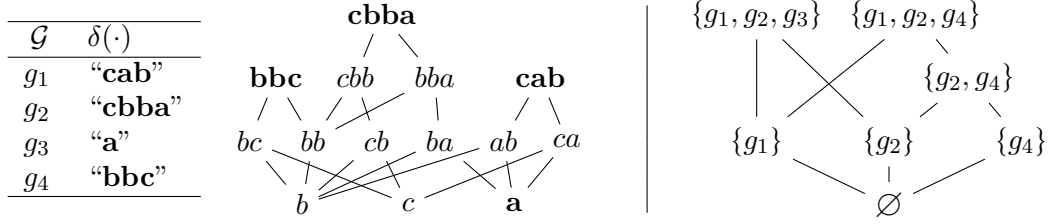


Figure 4. The table (left) represents the mapping function δ of the pattern setup considered in running example 3.3. The diagram (center) represents the set of non empty substrings in $\{a, b, c\}^+$ holding for at least one object in \mathcal{G} . The diagram (right) represents the poset of definable sets $(\mathbb{P}_{ext}, \sqsubseteq)$.

ciates to each objects in \mathcal{G} its word in the description space. For instance $\delta(g_1) = \text{“cab”}$. The diagram in the center of Fig. 4 depicts the Hasse Diagram of the poset $(\downarrow \delta[\mathcal{G}], \sqsubseteq)$ with $\delta[\mathcal{G}] = \{\text{“cab”}, \text{“cbba”}, \text{“a”}, \text{“bbc”}\}$. In other words, it depicts the set of descriptions $d \in \mathcal{D}$ holding for at least one object in \mathcal{G} . It is clear that the description “ca” holds for g_1 since “ca” \sqsubseteq “cab”. However, description “cb” does not hold for g_1 since “cb” is not a substring of “cab”. More generally, descriptions holding for g_i is the principal filter of $\delta(g_i)$ (i.e. $\downarrow \delta(g_i)$). For instance, the set of descriptions holding for g_1 is given by $\downarrow \delta(g_1) = \{\text{“a”}, \text{“ca”}, \text{“ab”}, \text{“cab”}\}$.

Example 3.4. Consider the *pattern setup* $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ presented in Fig. 5 (left). The set of objects is $\mathcal{G} = \{g_i\}_{1 \leq i \leq 4}$ and the description space $\underline{\mathcal{D}}$ is the powerset ordered by set inclusion $(\wp(\mathcal{M}), \sqsubseteq)$ (i.e. itemsets) with $\mathcal{M} = \{a, b, c\}$. Again descriptions holding for g_4 are all itemsets included in $\delta(g_4) = \{b, c\}$ (i.e. $\downarrow \delta(g_4) = \{\emptyset, \{b\}, \{c\}, \{b, c\}\}$).

3.1. On Extent and Cover Operators

We have seen that the relation “**realizes**” build a binary relation between objects and descriptions. Based on this binary relation, two key operators, namely **extent** and **cover** are derived (see Definition 3.5 and 3.6).

Definition 3.5. The **extent operator**, denoted by *ext*, is the operator that takes each description $d \in \mathcal{D}$ to the subset of objects in \mathcal{G} *realizing* it:

$$ext : \mathcal{D} \rightarrow \wp(\mathcal{G}), d \mapsto \{g \in \mathcal{G} \mid d \sqsubseteq \delta(g)\}$$

The size of $ext(d)$ is called the **support** of d , i.e. $support : d \mapsto |ext(d)|$.

Note 11. Please note that for any $S \subseteq \mathcal{D}$, we denote $ext[S] = \{ext(d) \mid d \in S\}$.

Definition 3.6. The **cover operator**, denoted by *cov*, takes each subset of objects $A \subseteq \mathcal{G}$ to the set of common descriptions in \mathcal{D} *covering* all of them:

$$cov : \wp(\mathcal{G}) \rightarrow \wp(\mathcal{D}), A \mapsto \delta[A]^\ell = \{d \in \mathcal{D} \mid (\forall g \in A) d \sqsubseteq \delta(g)\}$$

Example 3.7. Consider the pattern setup presented in Example 3.3. We have: $ext(\text{“bb”}) = \{g_2, g_4\}$ and $cov(\{g_2, g_4\}) = \downarrow \delta(g_2) \cap \downarrow \delta(g_4) = \{\text{“b”}, \text{“bb”}, \text{“c”}\}$.

Definition 3.8. A subset $A \subseteq \mathcal{G}$ is said to be:

- **Definable, Separable** or an **Extent** if $(\exists d \in \mathcal{D}) A = ext(d)$.
- **Coverable** if $cov(A) \neq \emptyset$.

The set of **definable sets** is then given by:

$$\mathbb{P}_{ext} = ext[\mathcal{D}] = \{ext(d) \mid d \in \mathcal{D}\}$$

Note 12. One should note that poset $(\mathbb{P}_{ext}, \subseteq)$ does form a subposet of $(\wp(\mathcal{G}), \subseteq)$, that is definable sets are naturally ordered by \subseteq . The set of *coverable sets* is naturally given by $\downarrow \mathbb{P}_{ext}$. In other words, any subset of a *coverable set* is *coverable*. Conversely, any superset of a *non coverable set* is a *non coverable set*.

Example 3.9. The poset of definable sets $(\mathbb{P}_{ext}, \subseteq)$ associated to the pattern setup presented in Example 3.3 is depicted in Fig. 4 (right). For example, it is clear that $\{g_2, g_4\}$ is definable since $ext("bb") = \{g_2, g_4\}$. However, there is no description which extent is exactly $\{g_1, g_2\}$, hence $\{g_1, g_2\}$ will be said **non-definable**. Still, $\{g_1, g_2\}$ is coverable since g_1 and g_2 share at least one common description (i.e. $cov(\{g_1, g_2\}) = \{“a”, “b”, “c”\} \neq \emptyset$). One should note also that $\{g_3, g_4\}$ is **non-coverable** since they share no common symbol and the empty string is excluded from the pattern space.

An important property arising directly from the definition of both *ext* and *cov* is given in Proposition 3.10. It tells that: on the one hand, the more restrictive is a description, the less it covers objects in the database. On the other hand, the bigger is a set of objects the less they share descriptions in common.

Proposition 3.10. *Operators ext and cov are order-reversing:*

$$(\forall c, d \in \mathcal{D}) \quad c \subseteq d \Rightarrow ext(d) \subseteq ext(c) \quad (\forall A, B \subseteq \mathcal{G}) \quad A \subseteq B \Rightarrow cov(B) \subseteq cov(A)$$

Proof. We have:

- (1) Let $A \subseteq B \subseteq \mathcal{G}$, let $d \in cov(B)$, thus $(\forall g \in B) d \subseteq \delta(g)$. Since $A \subseteq B$, we conclude that $(\forall g \in A) d \subseteq \delta(g)$ that is $d \in cov(A)$. Thus $cov(B) \subseteq cov(A)$.
- (2) Let $c, d \in \mathcal{D}$ such that $c \subseteq d$. Let $g \in ext(d)$, that is $d \subseteq \delta(g)$ thus $c \subseteq \delta(g)$; that is $g \in ext(c)$. We conclude that $ext(d) \subseteq ext(c)$.

This concludes the proof. □

Note 13. We have seen in Proposition 3.10 that mappings *ext* and *cov* are order reversing. Hence, one could think that (ext, cov) may form some Galois connection¹. However, it is not the case since *ext* associate an extent to one description while *cov* outputs a set of descriptions rather than one. In other words, these two mappings are not compatibles. Yet, we will see in next section that *ext* will be involved into a Galois connection when the considered pattern setup verifies additional properties (i.e. the pattern setup is a pattern structure Ganter and Kuznetsov (2001)). Mapping *cov* will also be involved in another Galois connection in Section 8.

Definition 3.11. For $c, d \in \mathcal{D}$, the **pattern implication** $c \rightarrow d$ holds if $ext(c) \subseteq ext(d)$. That is, every object realizing c realizes d . Dually, for $A, B \subseteq \mathcal{G}$, the **object implication** $A \rightarrow B$ holds if $cov(A) \subseteq cov(B)$. That is, every description covering all object in A covers also all objects in B .

We say that descriptions $c, d \in \mathcal{D}$ are **equivalent** if $c \rightarrow d$ and $d \rightarrow c$ and we have $ext(c) = ext(d)$. Dual definition can be given for the *equivalence between object sets*.

¹A Galois connection between two posets (P, \leq) and (Q, \sqsubseteq) is a pair (f, g) with $f : P \rightarrow Q$ and $g : Q \rightarrow P$ are order-reversing and both operators $f \circ g$ and $g \circ f$ are extensive.

Note 14. Please note that if $d \sqsubseteq c$ and since ext is an *order-reversing mapping*, we have $c \rightarrow d$. Regarding this observation, there is two types of implications between descriptions: (i) implications deduced directly from \sqsubseteq and (ii) implications that are dependent on the pattern setup. While the former implications are intrinsic to the description space, the latter are more important since they are those who enclose the knowledge hidden in the pattern setup.

Example 3.12. In the pattern setup presented in Example 3.3 and Fig. 4, we have $ext("bb") = \{g_2, g_4\}$ and $ext("c") = \{g_1, g_2, g_4\}$. Hence, we have $"bb" \rightarrow "c"$ or in other words in every string containing $"bb"$ in the pattern setup contains also $"c"$.

Proposition 3.13 gives characterizations of the set $ext[cov(A)]$ and $cov(ext(d))$ for $A \subseteq \mathcal{G}$ and $d \in \mathcal{D}$. This proposition will be useful later in this paper.

Proposition 3.13. For $A \subseteq \mathcal{G}$ and $d \in \mathcal{D}$:

$$\begin{aligned} ext[cov(A)] &= \{E \in \mathbb{P}_{ext} \mid A \subseteq E\} = \uparrow A \cap \mathbb{P}_{ext} \\ cov(ext(d)) &= \{c \in \mathcal{D} \mid ext(d) \subseteq ext(c)\} = \{c \in \mathcal{D} \mid d \rightarrow c\} \end{aligned}$$

Proof. We show the two equations separately:

(1) Let $E \subseteq \mathcal{G}$, we have:

$$\begin{aligned} E \in ext[cov(A)] &\Leftrightarrow (\exists d \in cov(A)) E = ext(d) \Leftrightarrow (\exists d \in \mathcal{D} \forall g \in A) d \sqsubseteq \delta(g) \\ &\Leftrightarrow (\exists d \in \mathcal{D}) A \subseteq ext(d) = E \Leftrightarrow E \in \mathbb{P}_{ext} \cap \uparrow A \end{aligned}$$

We conclude that $ext[cov(A)] = \{E \in \mathbb{P}_{ext} \mid A \subseteq E\} = \uparrow A \cap \mathbb{P}_{ext}$

(2) Let $c \in \mathcal{D}$, we have:

$$\begin{aligned} c \in cov(ext(d)) &\Leftrightarrow (\forall g \in ext(d)) c \sqsubseteq \delta(g) \Leftrightarrow (\forall g \in ext(d)) g \in ext(c) \\ &\Leftrightarrow ext(d) \subseteq ext(c) \end{aligned}$$

Thus $cov(ext(d)) = \{c \in \mathcal{D} \mid ext(d) \subseteq ext(c)\} = \{c \in \mathcal{D} \mid c \rightarrow d\}$. □

Example 3.14. Consider the pattern setup presented in Example 3.3 and its associated $(\mathbb{P}_{ext}, \sqsubseteq)$ depicted in Fig. 4 (right). We have: $ext("bb") = \{g_2, g_4\}$ and $cov(\{g_2, g_4\}) = \{"b", "bb", "c"\}$. Hence:

- $ext[cov(\{g_2, g_4\})] = \{ext("b"), ext("bb"), ext("c")\} = \{\{g_1, g_2, g_4\}, \{g_2, g_4\}\}$.
- $cov(ext("bb")) = cov(\{g_2, g_4\}) = \{"b", "bb", "c"\}$.

3.2. A Minimal Representation of a Pattern Setup

An important notion analogous to what is called *representation context* in pattern structures (see Ganter and Kuznetsov (2001); Buzmakov, Kuznetsov, and Napoli (2015)) is introduced in Theorem 3.17. Technically, such a representation does not provide a practical way to explore definable sets of an arbitrary pattern setups, but helps to simulate definable sets search space of a pattern setup independently from the description space. Before introducing the Theorem, we present Proposition 3.15.

Proposition 3.15. *Let \mathcal{G} be a non empty finite set and let $S \subseteq \wp(\mathcal{G})$, we have*

$$\exists \mathbb{P} \text{ a pattern setup such that } S = \mathbb{P}_{ext} \iff (\forall g \in \mathcal{G}) \bigcap (\uparrow g \cap S) \in S$$

Proof. Before showing the equivalence let us prove the following property:

$$(\forall g \in \mathcal{G}) \bigcap (\uparrow g \cap \mathbb{P}_{ext}) = ext(\delta(g)) \in \mathbb{P}_{ext} \quad (1)$$

Recall that $ext[cov(\{g\})] = \uparrow g \cap \mathbb{P}_{ext}$ (See proposition 3.13). Let $g \in \mathcal{G}$, we have $\delta(g) \in cov(\{g\})$, thus $ext(\delta(g)) \in ext[cov(\{g\})]$. Let us show that $ext(\delta(g))$ is a lower bound of $ext[cov(\{g\})]$. We have: $cov(\{g\}) = \{d \in \mathcal{D} \mid d \sqsubseteq \delta(g)\}$. Thus, $\forall d \in cov(\{g\}) : d \sqsubseteq \delta(g)$. Since ext is an order reversing operator, we obtain: $\forall A \in ext[cov(\{g\})] : ext(\delta(g)) \subseteq A$. Thus, $ext(\delta(g))$ is the smallest element of $ext[cov(\{g\})]$. That is $\bigcap (\uparrow g \cap \mathbb{P}_{ext}) = ext(\delta(g))$.

We show now the two implications independently:

(\Rightarrow) Let $S = \mathbb{P}_{ext}$ for some pattern setup \mathbb{P} . Using eq. (1), $\bigcap (\uparrow g \cap S) \in S$.

(\Leftarrow) Let $S \subseteq \wp(\mathcal{G})$ for which $\forall g \in \mathcal{G}$ we have $\bigcap (\uparrow g \cap S) \in S$. Let us now define the following pattern setup:

$$\mathbb{P} = \left(\mathcal{G}, (S, \supseteq), \delta : g \mapsto \bigcap (\uparrow g \cap S) \right)$$

Let $A \in S$ be a description, we have: $ext(A) = \{g \in \mathcal{G} \mid \bigcap (\uparrow g \cap S) \subseteq A\}$. Let us show that $ext(A) = A$ by showing double inclusion: **(1)** Let $g \in A$, thus $A \in (\uparrow g \cap S)$. It follows that $\bigcap (\uparrow g \cap S) \subseteq A$. We conclude that $g \in ext(A)$. Therefor $A \subseteq ext(A)$. **(2)** Let $g \in ext(A)$, thus $\bigcap (\uparrow g \cap S) \subseteq A$. Since $\forall B \in \uparrow g \cap S$ we have $g \in B$, we have $g \in \bigcap (\uparrow g \cap S)$, that is $g \in A$. We conclude that $ext(A) \subseteq A$. Both inclusion leads us to have $(\forall A \in S) ext(A) = A$. We conclude that $ext[S] = S$. In other words, $\mathbb{P}_{ext} = S$.

This concludes the proof. \square

Example 3.16. Proposition 3.15 tells that not all families of subsets of \mathcal{G} could be seen as a set of extents of some pattern setup. Consider the poset depicted in Fig. 2 (4) where $\mathcal{G} = \{g_1, g_2, g_3, g_4\}$ and $S = \{\{g_1, g_2\}, \{g_1, g_3\}, \{g_1, g_2, g_3\}, \{g_1, g_2, g_3, g_4\}\}$ can never be seen as a \mathbb{P}_{ext} for some pattern setup \mathbb{P} . Indeed, $\uparrow g_1 \cap S = \{\{g_1, g_2\}, \{g_1, g_3\}\}$ whose intersection is not in S .

Theorem 3.17. *For any pattern setup \mathbb{P} , the pattern setup $\mathbb{R}(\mathbb{P})$ given by*

$$\mathbb{R}(\mathbb{P}) = \left(\mathcal{G}, (\mathbb{P}_{ext}, \supseteq), g \mapsto \bigcap (\uparrow g \cap \mathbb{P}_{ext}) \right)$$

*is called the **minimal representation** of \mathbb{P} and we have $\mathbb{R}(\mathbb{P})_{ext} = \mathbb{P}_{ext}$.*

Proof. Theorem 3.17 is a corollary of Proposition 3.15. Indeed, the pattern setup $\mathbb{R}(\mathbb{P})$ is the same as the one built in the proof of Proposition 3.15 (\Leftarrow) since \mathbb{P}_{ext} is a set system verifying the property $(\forall g \in \mathcal{G}) \bigcap (\uparrow g \cap \mathbb{P}_{ext}) \in \mathbb{P}_{ext}$ (i.e. implication (\Rightarrow)). We have $\mathbb{R}(\mathbb{P})_{ext} = \mathbb{P}_{ext}$. Moreover, this representation is said to be minimal since any proper subset of $(\mathbb{P}_{ext}, \supseteq)$ will drop at least one definable set. \square

4. Pattern Structures

Pattern structures were introduced in Ganter and Kuznetsov (2001). They require that every set of objects has a greatest common description (least general generalization). A formal definition is given in Definition 4.1. Pattern structures provide a very strong tool to formalize a large class of pattern languages (Kuznetsov 2009). For instance, pattern setups over the language of itemsets (Ganter and Wille 1999), intervals (Kaytoue, Kuznetsov, and Napoli 2011), convex polygons (Belfodil et al. 2017), sequence sets (Buzmakov et al. 2016)², and graph sets²(Kuznetsov 1999) are all pattern structures.

Definition 4.1. A pattern setup $\mathbb{P} = (\mathcal{G}, \mathcal{D}, \delta)$ is said to be a **pattern structure** iff every subset of objects has a greatest common description. Formally:

$$(\forall S \subseteq \delta[\mathcal{G}]) \ S \text{ does have a meet } \bigsqcap S$$

Example 4.2. The pattern setup presented in Example 3.4 and Fig. 5 is a pattern structure. Indeed, since the description space is the powerset lattice $(\wp(\{a, b, c\}), \subseteq)$ (i.e a complete lattice) then every subsets $S \subseteq \delta[\mathcal{G}] \subseteq \mathcal{D}$ does have a meet which is the set intersection $\bigcap S$.

However, the pattern setup \mathbb{P} presented in Example 3.3 and Fig. 4 is not a pattern structure. Indeed, the set of common descriptions $cov(\{g_2, g_4\}) = \{“b”, “bb”, “c”\}$ does not have a maximum (i.e. $\{\delta(g_2), \delta(g_4)\}$ does not have a meet) since $\{“b”, “bb”, “c”\}$ has two maximal elements.

One can define a new operator, namely the **intent**, in a pattern structure thanks to the existence of the meet.

Definition 4.3. The **intent operator**, denoted by int , is the operator that takes each subset of objects $A \subseteq \mathcal{G}$ to the greatest common description in \mathcal{D} covering them (i.e. the maximum of $cov(A)$). Formally:

$$int : \wp(\mathcal{G}) \rightarrow \mathcal{D}, A \mapsto inf \delta[A] = \bigsqcap \delta[A]$$

Note 15. In a pattern structure, the pair of operators (ext, int) forms a *Galois connection* between posets $(\wp(\mathcal{G}), \subseteq)$ and $(\mathcal{D}, \sqsubseteq)$. Thus, $ext \circ int$ and $int \circ ext$ form *closure operators* (cf. Proposition 8 in Ganter and Wille (1999) book) on the two posets respectively. Thanks to this Galois Connection, one can define a complete lattice based on the the closed elements.

Definition 4.4. Let $\mathbb{P} = (\mathcal{G}, \mathcal{D}, \delta)$ be a *pattern structure*. The **(pattern) concept lattice** associated to \mathbb{P} is the *complete lattice* denoted by $\mathfrak{B}(\mathbb{P}) = (\mathfrak{B}(\mathbb{P}), \leq)$. Elements of $\mathfrak{B}(\mathbb{P})$ are called **(pattern) concepts** and are given by:

$$(A, d) \in \wp(\mathcal{G}) \times \mathcal{D} \quad s.t. \quad A = ext(d) \text{ and } d = int(A)$$

The concepts are ordered by \leq as follows: $(A_1, d_1) \leq (A_2, d_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow d_2 \sqsubseteq d_1$.

Note 16. Two complete lattices isomorphic to the concept lattice can be derived:

²Sequence and graphs patterns will be discussed in the next section, since they do not induce pattern structures directly, but the sets of incomparable patterns do.

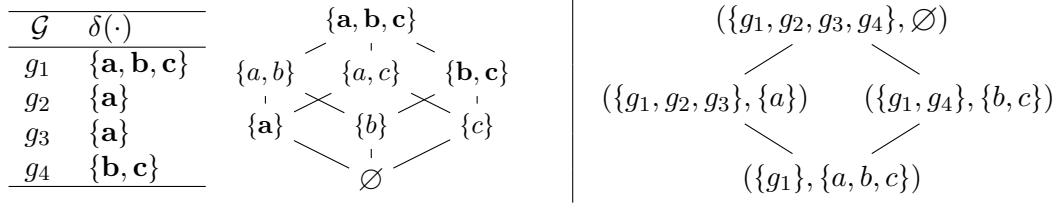


Figure 5. The table (left) represents the pattern setup $\mathbb{P} = (\mathcal{G}, \mathcal{D}, \delta)$ with $\mathcal{G} = \{g_i\}_{1 \leq i \leq 4}$, $\mathcal{D} = (\wp(\{a, b, c\}), \subseteq)$ depicted by the Hasse Diagram (center) and δ maps an object to its itemset. The diagram (right) depicts the concept lattice $\underline{\mathcal{B}}(\mathbb{P})$.

- (1) The poset of *definable sets* $(\mathbb{P}_{ext}, \subseteq)$ which on a \cap -structures (i.e. *Moore family, closure system*) in the powerset lattice $(\wp(\mathcal{G}), \subseteq)$. Note that *definable sets* are the fixpoints of the closure operator $ext \circ int$.
- (2) The poset of *closed patterns* $(\mathcal{D}_\delta, \supseteq)$ with $\mathcal{D}_\delta = int[\wp(\mathcal{G})] = \{\bigcap \delta[A] \mid A \subseteq \mathcal{G}\}$ is a complete lattice. Elements of \mathcal{D}_δ are called *closed patterns* since they are fixpoints of the closure operator $int \circ ext$.

Another important remark about the closure operator $ext \circ int$ is that it takes to a subset of object $A \subseteq \mathcal{G}$ the smallest definable set $E \in \mathbb{P}_{ext}$ enclosing it. Formally:

Proposition 4.5. *Let $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \subseteq), \delta)$ be a pattern structure, we have:*

$$ext \circ int : \wp(\mathcal{G}) \rightarrow \wp(\mathcal{G}), A \mapsto \bigcap \{E \in \mathbb{P}_{ext} \mid A \subseteq E\} = \bigcap (\uparrow A \cap \mathbb{P}_{ext})$$

Proof. This result is straightforward from the fact that $ext \circ int$ is a closure operator. Indeed, according to Theorem 1 in Ganter and Wille (1999) (page 8), we have $\mathbb{P}_{ext} = \{ext \circ int(A) \mid A \subseteq \mathcal{G}\}$ is closure system. By application of the theorem we have: $ext \circ int : A \mapsto \bigcap \{E \in \mathbb{P}_{ext} \mid A \subseteq E\}$. \square

Example 4.6. Consider again the *pattern structure* $\mathbb{P} = (\mathcal{G}, (\wp(\{a, b, c\}), \subseteq), \delta)$ presented in Fig. 5. Since the meet is the set intersection we have:

$$int(\{g_1, g_4\}) = \delta(g_1) \cap \delta(g_4) = \{a, b, c\} \cap \{b, c\} = \{b, c\}$$

The concept lattice $\underline{\mathcal{B}}(\mathbb{P})$ is depicted in Fig. 5 (right). One should note the set of definable sets $(\mathbb{P}_{ext}, \subseteq)$ can be deduced directly from the concept lattice by taking extents of the pattern concepts. It is important to highlight the fact that that pattern structure \mathbb{P} is derived from the formal context $\mathbb{K} = (\mathcal{G}, \mathcal{M}, \mathcal{I})$ presented in Fig. 3 where δ is given by $\delta : g \mapsto \{m \in \mathcal{M} \mid g \mathcal{I} m\}$.

Definition 4.1 follows Lumpe and Schmidt (2015). The original equivalent definition of the pattern structure (Ganter and Kuznetsov 2001) requires that the description space must be a complete lattice. Theorem 4.7 builds a bridge between meet-semilattices and pattern structures over finite set of objects or more generally between pattern structures and complete lattices.

Theorem 4.7. Let $\underline{\mathcal{D}} = (\mathcal{D}, \sqsubseteq)$ be a poset, the following properties are equivalent:

- For any finite set $\mathcal{G} \neq \emptyset$ and any $\delta \in \mathcal{D}^{\mathcal{G}}$, $(\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is a pattern structure (where $\mathcal{D}^{\mathcal{G}}$ denotes the set of all mappings $\delta : \mathcal{G} \rightarrow \mathcal{D}$).
- $\underline{\mathcal{D}}$ is an upper-bounded meet-semilattice (i.e. \emptyset has also a meet).

For an arbitrary set \mathcal{G} , the following properties are equivalent:

- For any set $\mathcal{G} \neq \emptyset$ and any $\delta \in \mathcal{D}^{\mathcal{G}}$, $(\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is a pattern structure.
- $\underline{\mathcal{D}}$ is a complete lattice.

Proof. Let us show both implications for a finite \mathcal{G} :

\Rightarrow The empty set has a meet in $\underline{\mathcal{D}}$ since $\delta[\emptyset] = \emptyset$ has a meet. Thus $\underline{\mathcal{D}}$ has a top element $\top = \bigsqcup \mathcal{D} = \bigsqcup \emptyset$. Moreover, let $S \subseteq \mathcal{D}$ be a finite set, one can build a finite set \mathcal{G} such that $\delta[\mathcal{G}] = S$. Since \mathbb{P} is a pattern structure then $S = \delta[\mathcal{G}]$ has a meet. We conclude that $\underline{\mathcal{D}}$ is an upper-bounded meet-semilattice.

\Leftarrow Let $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ be a pattern setup. Any subset of $\delta[\mathcal{G}]$ is finite subset of \mathcal{D} and thus has a meet (including the \emptyset since \mathcal{D} has its top element).

Let us now consider the case of arbitrary set \mathcal{G} :

\Rightarrow Let $S \subseteq \mathcal{D}$, one can build \mathcal{G} such that $\delta[\mathcal{G}] = S$. Since \mathbb{P} is a pattern structure then $\delta[\mathcal{G}] = S$ has a meet. We conclude that $\underline{\mathcal{D}}$ is a complete lattice.

\Leftarrow Let $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ be a pattern setup. Any subset of $\delta[\mathcal{G}]$ is a subset of \mathcal{D} and thus has a meet (including the \emptyset since \mathcal{D} has its top element).

This concludes the proof. \square

The state-of-the-art abounds with examples of descriptions spaces that are complete lattices that someone can use to build pattern structures:

- *Itemset pattern structure (Ganter and Wille 1999)*. The description space is the Boolean lattice $(\wp(M), \subseteq)$ where M is a non empty finite set of attributes.
- *Interval pattern structure (Kaytoue, Kuznetsov, and Napoli 2011)*. The description space is the complete lattice $(\mathcal{C}(\mathbb{R})^m, \supseteq)^3$, where $\mathcal{C}(\mathbb{R})^m$ represents the set of all possible axis-parallel m -dimensional hyperrectangles in \mathbb{R}^m (m is the number of attributes) and \sqsubseteq represents the hyperrectangle inclusion.
- *Convex sets pattern structure (Belfodil et al. 2017)*. The description space is the complete lattice of all convex sets in \mathbb{R}^m ordered by inclusion $(\mathcal{C}(\mathbb{R}^m), \supseteq)$.
- *Partition pattern structure (Codoceo and Napoli 2014)*. The description space is the complete lattice of all partitions $(\mathcal{B}(E), \sqsubseteq)$ of some finite set E . The order \sqsubseteq is *finer-than* order relation between partitions. That is for $P_1, P_2 \in \mathcal{B}(E)$ two partitions, $P_1 \sqsubseteq P_2$ if and only $(\forall E_1 \in P_1 \exists E_2 \in P_2) E_1 \subseteq E_2$.

Note 17. Before finishing this section, let us highlight some key differences between arbitrary pattern setups and pattern structures. It is clear that the main difference is the fact that the greatest common description does not necessarily exist for any subsets of objects in an arbitrary pattern setup. This implies that the set of definable sets $(\mathbb{P}_{ext}, \subseteq)$ is not necessarily closed under intersection in an arbitrary pattern setup as shown in Fig. 4. One should also note that, conversely to pattern setups where some subsets can be even non-coverable (see Example 3.9), In Pattern Structures, all subsets $A \subseteq \mathcal{G}$ are coverables since $cov(A) \supseteq cov(\mathcal{G}) \neq \emptyset$. In fact, \mathcal{G} is always definable since $ext(\bigsqcup \delta[\mathcal{G}]) = \mathcal{G}$.

³ $\mathcal{C}(E)$ is the set of convex subsets of E . The set $\mathcal{C}(\mathbb{R})$ is the set of all possible intervals of \mathbb{R} , $\mathcal{C}(\mathbb{R})^m$ is then the set of all axis-parallel m -dimensional hyperrectangles.

5. From Closed Patterns to Support-Closed Patterns

In pattern mining, another notion of closedness is generally considered (Yan, Han, and Afshar 2003; Wang and Han 2004). Definition 5.1 defines formally such a notion dubbed *support-closedness* by Boley et al. (2010).

Definition 5.1. A description d is said to be **support-closed** in a pattern setup iff:

$$(\forall c \in \mathcal{D}) \ d \sqsubset c \implies \text{ext}(c) \sqsubset \text{ext}(d).$$

We will see below that this notion is linked to maximal common descriptions.

5.1. On Maximal Common Descriptions

In pattern structures, support-closed patterns coincide exactly with closed descriptions (i.e. fixpoints of $\text{int} \circ \text{ext}$) since int takes a subset of objects to the greatest common description. However, when we consider an arbitrary pattern setup, such a maximum common description may not exist (see Example 4.2). One straightforward generalization is to associate to a subset of object the set of its maximal common descriptions (see Definition 5.2). Proposition 5.3 builds then a bridge between support-closed patterns and maximal common descriptions.

Definition 5.2. The set of **maximal covering (common) descriptions** of a subset $A \subseteq \mathcal{G}$, denoted by $\text{cov}^*(A)$, is given by:

$$\text{cov}^* : A \mapsto \max(\text{cov}(A)) = \max(\delta[A]^\ell)$$

Proposition 5.3. A description $d \in \mathcal{D}$ is support-closed iff $(\exists A \subseteq \mathcal{G}) \ d \in \text{cov}^*(A)$. The set of all support-closed descriptions is given by: $\mathcal{D}^* = \bigcup_{A \subseteq \mathcal{G}} \text{cov}^*(A)$.

Proof. We prove the two implications:

- (\implies) Let $d \in \mathcal{D}$ be a support-closed description and let $A = \text{ext}(d)$. Hence, according to proposition 3.13 we have $d \in \text{cov}(A)$. Let us show now that $d \in \text{cov}^*(A)$. Suppose that $d \notin \text{cov}^*(A)$ that is $\uparrow d \cap \text{cov}(A) \neq \{d\}$. Since $d \in \text{cov}(A)$, there is then at least $c \in \text{cov}(A)$ such that $d \sqsubset c$. Thus, in one hand and according to proposition 3.10, $\text{ext}(c) \subseteq \text{ext}(d)$. And since $c \in \text{cov}(A)$, according to proposition 3.13, $\text{ext}(c) \supseteq \text{ext}(d)$. Thus $\text{ext}(c) = \text{ext}(d)$. This is contradictory with the fact that d is support-closed ($\exists c \in \mathcal{D}$ s.t. $d \sqsubset c$ and $\text{ext}(c) = \text{ext}(d)$). Therefore, $(\exists A \subseteq \mathcal{G}) \ d \in \text{cov}^*(A)$.
- (\impliedby) Suppose that $\exists A \subseteq \mathcal{G}$ s.t. $d \in \max(\text{cov}(A))$. According to proposition 3.13 and since $d \in \text{cov}(A)$, we have $A \subseteq \text{ext}(d)$. Let now be $c \in \mathcal{D}$ such that $d \sqsubset c$, we have $c \notin \text{cov}(A)$ since d is maximal in $\text{cov}(A)$. According to proposition 3.10 we have $\text{ext}(c) \subseteq \text{ext}(d)$. Moreover, using proposition 3.13 and since $c \notin \text{cov}(A)$ we have $A \not\subseteq \text{ext}(c)$. Since $A \subseteq \text{ext}(d)$ then $\text{ext}(c) \neq \text{ext}(d)$. Thus $\forall c \in \mathcal{D}$ such that $d \sqsubset c$ we have $\text{ext}(c) \sqsubset \text{ext}(d)$; that is d is support-closed.

The formula of \mathcal{D}^* is deduced directly. Please notice also that if there exists A s.t. $d \in \text{cov}^*(A)$ then $d \in \text{cov}^*(\text{ext}(d))$ (use (\impliedby) then (\implies)). Hence, $d \in \mathcal{D}$ is support-closed iff $d \in \text{cov}^*(\text{ext}(d))$. \square

Example 5.4. Reconsider Example 3.3, we have $cov(\{g_2, g_4\}) = \{\text{"b"}, \text{"bb"}, \text{"c"}\}$. Hence, the maximal covering ones are given by $cov^*(\{g_2, g_4\}) = \{\text{"bb"}, \text{"c"}\}$.

Note 18. In a pattern structure we have: $cov^* : \wp(\mathcal{G}) \rightarrow \wp(\mathcal{D}), A \mapsto \{int(A)\}$.

5.2. On Upper-Approximation Extents

Going back to pattern structures, the closure operator $ext \circ int$ takes any subset $A \subseteq \mathcal{G}$ to the smallest definable set $ext(int(A))$ enclosing it as stated by Proposition 4.5. This fact is used to enumerate all definable sets via the closure operator (see Kuznetsov (1993, 1999)). From Rough Set Theory (Pawlak 1982) perspective, the set $ext(int(A))$ can be seen as the upper approximation of an arbitrary and potentially non definable set A in \mathbb{P}_{ext} . However, when it comes to an arbitrary pattern setup, a non-definable set A may have many *minimal* definable sets enclosing it or no one if it is non-coverable (see Example 5.6). Definition 5.5 formalizes this second generalization.

Definition 5.5. The **set of upper-approximation extents** of a subset $A \subseteq \mathcal{G}$, denoted by \overline{A} , is given by the set of minimal definable sets in \mathbb{P}_{ext} enclosing A :

$$\overline{A} = \min(\{E \in \mathbb{P}_{ext} \mid A \subseteq E\}) = \min(\uparrow A \cap \mathbb{P}_{ext}).$$

Example 5.6. Consider Example 3.3, the upper approximations of subset $A = \{g_2, g_4\}$ is given by $\overline{A} = \{A\}$ since A is definable. For the set $B = \{g_1, g_2\}$, we have $\overline{B} = \{\{g_1, g_2, g_3\}, \{g_1, g_2, g_4\}\}$ that is B has two upper-approximation extents. For $C = \{g_3, g_4\}$, it is clear that there is no definable set in \mathbb{P}_{ext} enclosing C (see Fig. 4 (right)), thus $\overline{C} = \emptyset$.

Note 19. According to Proposition 3.13, we have $(\forall A \subseteq \mathcal{G}) \overline{A} = \min(ext[cov(A)])$. Moreover, in a pattern structure, $\overline{A} = \{ext(int(A))\}$ for all $A \subseteq \mathcal{G}$.

5.3. Linking Upper-Approximation Extents to Support-Closed Patterns.

We have seen before that on the one hand cov^* operator is somehow a generalization of pattern structure *int operator* in an arbitrary pattern setup and on the other hand, *upper-approximation extents* operator is a generalization of pattern structure closure operator $ext \circ int$. Indeed, in a pattern structure we have for $A \subseteq \mathcal{G}$, $cov^*(A) = \{int(A)\}$ and $\overline{A} = \{ext(int(A))\}$. That is: $\overline{A} = ext[cov^*(A)]$. One judicious question is that, does this property still hold for an arbitrary pattern setup? Let us analyze the following example.

Example 5.7. Reconsider Example 3.3 and the definable set $A = \{g_2, g_4\}$. We have $cov(A) = \{\text{"b"}, \text{"bb"}, \text{"c"}\}$ and $cov^*(A) = \{\text{"bb"}, \text{"c"}\}$. Thus on the one hand, we have:

$$ext[cov^*(A)] = \{ext(\text{"bb"}), ext(\text{"c"})\} = \{\{g_2, g_4\}, \{g_1, g_2, g_4\}\}$$

but on the other hand, since A is definable, we have $\overline{A} = \{\{g_2, g_4\}\}$.

Hence, according to Example 5.7, it is clear that the property $\overline{A} = ext[cov^*(A)]$ does not hold in a pattern setup. In fact, things can go even worse when the description space is an *arbitrary infinite poset*. Let us analyze for that a second Example:

Example 5.8. Let be the poset $(\mathcal{D}, \sqsubseteq)$ presented in Fig. 6 where:

- $\mathcal{D} = \{a, b\} \cup \{c_i \mid i \in \mathbb{N}\}$,
- $(\forall i \in \mathbb{N}) c_i \sqsubseteq c_{i+1}, c_i \sqsubseteq a$ and $c_i \sqsubseteq b$.

Let be the pattern setup $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \sqsubseteq), \delta)$ such that $\mathcal{G} = \{g_1, g_2\}$, $\delta(g_1) = a$ and $\delta(g_2) = b$. It is clear that $\text{cov}(\{g_1, g_2\}) = \{c_i \mid i \in \mathbb{N}\}$. Yet, $\text{cov}^*(\{g_1, g_2\}) = \emptyset$ since $\text{cov}(\{g_1, g_2\})$ is an infinitely ascending chain. Therefore, there is no maximal common description covering both g_1 and g_2 . Thus, given $A \subseteq \mathcal{G}$ it seems that there is no link between $\text{cov}^*(A)$ and \bar{A} in an arbitrary pattern setup $(\mathcal{G}, \mathcal{D}, \delta)$. In fact, this even mean that considering only maximal common descriptions to look for all possible definable sets is totally a wrong idea since maximal covering descriptions do not hold all the information about definable sets. Indeed, while $\text{ext}[\mathcal{D}] = \{\{g_1\}, \{g_2\}, \{g_1, g_2\}\}$, the extents obtained from the set of support-closed patterns \mathcal{D}^* is given by:

$$\text{ext}[\mathcal{D}^*] = \{\{g_1\}, \{g_2\}\} \subsetneq \text{ext}[\mathcal{D}]$$

Going back to the case of pattern structures, there is a strong link between *int* and *cov*. Indeed, $(\forall A \subseteq \mathcal{G}) \text{cov}(A) = \downarrow \text{int}(A)$. In other words, knowing the *intent* of a subset of objects allows us to know every single common description to all objects in A . If we had want to generalize such a property for an arbitrary pattern setup we would have expected: $(\forall A \subseteq \mathcal{G}) \text{cov}(A) = \downarrow \text{cov}^*(A)$. In the sense that knowing maximal covering descriptions allows to deduce all the covering ones. Such a property does not hold for any pattern setup as shown in Example 5.8. This property is directly linked to what is called *multilattices* which we revisit in the next section.

6. Multilattices

The term **multilattice** was introduced for the first time by Benado (1955). This notion have not received much interest for a long period, but have been unearthed and revisited in the beginning of the 21st century by Martínez et al. (2001); Cordero et al. (2004); Martínez et al. (2005) for other purposes. We will start here by presenting multilattices following Martínez's et al. We will then understand the main difference between Benado's multilattices and (Martínez's et al.) multilattices afterward.

Before giving the formal definition of multilattices, we start by defining the notion of multi-infimum and multi-supremum. In the following section (P, \leq) denotes an arbitrary poset and $S \subseteq P$ denotes an arbitrary subset.

Definition 6.1. A **multi-infimum** (resp. **multi-supremum**) of S is a maximal (resp. minimal) element of S^ℓ (resp. S^u). The set of multi-infima (resp. multi-suprema) of S is denoted by $\text{minf}(S)$ (resp. $\text{msup}(S)$) and:

$$\text{minf}(S) = \max(S^\ell) \quad \text{msup}(S) = \min(S^u)$$

We say then that:

- S has all its **multi-infima** iff: $S^\ell = \downarrow \max(S^\ell) = \downarrow \text{minf}(S)$.
- S has all its **multi-suprema** iff: $S^u = \uparrow \min(S^u) = \uparrow \text{msup}(S)$.

Multilattices, as their names imply, are related in their definition with lattices. Simply put, multilattices are a relaxation of lattices where rather than demanding that the set of lower (resp. upper) bounds of each nonempty finite subset has its

infimum (resp. supremum), multilattices demand that the set of lower (resp. upper) bounds of each nonempty finite subset has all its multi-infima (resp. multi-suprema).

Definition 6.2. A poset (P, \leq) is said to be:

- A **meet-multisemilattice** if for all nonempty finite subsets $S \subseteq P$, S has all its mutli-infima.
- A **join-multisemilattice** if for all nonempty finite subsets $S \subseteq P$, S has all its mutli-suprema.
- A **multilattice** if it is both a meet-multisemilattice and a join-multisemilattice.
- A **complete meet-multisemilattice** if for all subsets $S \subseteq P$, S has all its mutli-infima.
- A **complete join-multisemilattice** if for all subsets $S \subseteq P$, S has all its mutli-suprema.
- A **complete multilattice** if it is both a complete meet-multisemilattice and a complete join-multisemilattice.

It is clear that all finite posets, or more generally finite-chain posets, are complete multilattices. More precisely, posets having the ascending (resp. descending) chain condition are complete join-multisemilattices (resp. meet-multisemilattices). It is also clear that all (complete) (semi)lattices are (complete) multi(semi)lattices since requiring that a subset $S \subseteq P$ to have all its multi-infima is weaker than requiring it to have an infimum. One should note also that since $\emptyset^\ell = \emptyset^u = P$ then we have:

- If (P, \leq) is a complete meet-multisemilattice then $P = \downarrow \max(P)$.
- If (P, \leq) is a complete join-multisemilattice then $P = \uparrow \min(P)$.

When we compare with lattices (cf. note 8), two questions straightforwardly raise:

- Are pair of elements the building blocks of a multilattice?
- Are all complete semimultilattices complete multilattices?

The followings sections answers negatively to these both questions.

6.1. On Benado's Multilattices

As said before, *multilattices* were introduced for the first time by Benado (1955). However, Benado defined multilattices as follow: a poset (P, \leq) is said to be multilattice if and only if all pairs of elements have all their multi-infima and all their multi-suprema. Formally:

$$\forall x, y \in P : \{x, y\}^\ell = \downarrow \minf(\{x, y\}) \text{ and } \{x, y\}^u = \uparrow \msup(\{x, y\}) \quad (2)$$

We will call here posets verifying such a property **Benado's multilattices**. As clearly explained in Martínez et al. (2005), such a property is not sufficient to have all non empty finite subsets have their multi-infima and multi-suprema. This is in contrast to lattices where it suffices to have a meet and join for all subsets of two elements to have the meet and join for all non empty finite subsets. For instance, one could verify that the poset depicted in Fig. 7 is a Benado's multilattice. Yet, it is not a multilattice according to Definition 6.2. For instance, the set $\{a, b, c\}$ does not have all its multi-infima. Indeed, $\{a, b, c\}^\ell = \{abc_i \mid i \in \mathbb{N}\}$, however, $\max(\{a, b, c\}^\ell) = \emptyset$. It follows that $\{a, b, c\}^\ell \neq \downarrow \max(\{a, b, c\}^\ell)$.

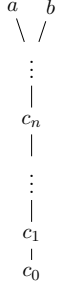


Figure 6. A complete join-multisemilattice but not a meet-multisemilattice

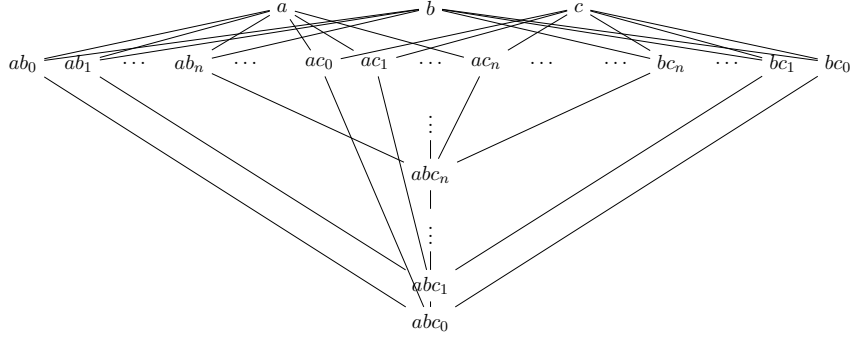


Figure 7. For all x, y in this poset, $\{x, y\}$ has all its multi-infima. That is, this poset is a multistructure following Benado (1955), however it is not a multilattice following our definition.

6.2. On Complete Multisemilattices

We have seen that all complete semilattices are complete lattices. However, this property no longer holds for complete multisemilattices. In fact one could have complete meet-multisemilattice that is even not a join-multisemilattice and *vice versa*. For instance, Fig. 6 depicts a complete join-multisemilattice that is even not a meet-multisemilattice. For instance, the set of lower bounds $\{a, b\}^\ell = \{c_i \mid i \in \mathbb{N}\}$ is an infinitely ascending chain and thus $\max(\{a, b\}^\ell) = \emptyset$. Thus, $\{a, b\}^\ell \neq \downarrow \max(\{a, b\}^\ell)$. In other words, $\{a, b\}$ does not have all its multi-infima.

6.3. On chain-completeness and complete multilattices

Complete multilattices are linked chain-complete posets. Before giving this relationship, let us recall the definition of chain-completeness.

Definition 6.3. A poset (P, \leq) is said to be:

- **Chain-complete** if all chains in P , including \emptyset , has its join.
- **Dually chain-complete** if all chains in P , including \emptyset , has its meet.
- **Doubly chain-complete** if it is chain-complete and dually chain-complete.

Since the empty set matches the definition of a chain-complete, all chain-complete posets are bounded. An important theorem linking complete lattices to chain-completeness is given in Theorem 6.4.

Theorem 6.4 (Theorem 3.24 from Roman (2008)). *A lattice (P, \leq) is a complete lattice if and only if it is chain-complete.*

One straightforward question is what is the relationship between complete multilattices and chain-complete posets. The answer is given in Theorem 6.5.

Theorem 6.5. *Under Axiom of Choice (AC) assumption⁴, we have:*

- *All chain-complete posets are complete meet-multisemilattice.*
- *All dually chain-complete posets are complete join-multisemilattice.*
- *All doubly chain-complete posets are complete multilattices.*

⁴I am grateful to JOZEF PÓCS for attracting my attention to Zorn's Lemma.

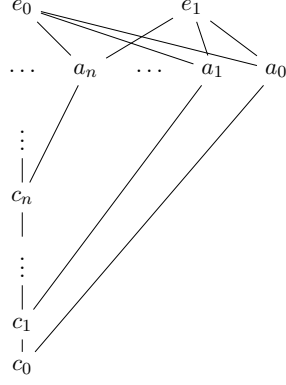


Figure 8. A complete multilattice that is not chain-complete

Proof. Before proving the theorem, we attract the reader to **Zorn's Lemma**. This lemma states that if every chain in a poset (P, \leq) has an upper-bound, then (P, \leq) has a maximal element. Formally:

$$(\forall C \in \mathcal{C}(P)) \ C^u \neq \emptyset \implies \max(P) \neq \emptyset$$

A stronger statement, yet equivalent, of Zorn's Lemma say even that:

$$(\forall C \in \mathcal{C}(P)) \ C^u \neq \emptyset \implies P = \downarrow \max(P) \tag{3}$$

Zorn's Lemma need to be considered as an **axiom** since it is equivalent to the well-known **axiom of choice (AC)**.

Let be now a chain-complete poset (P, \leq) , we need to show that (P, \leq) is a complete meet-multisemilattice. Let $S \subseteq P$ be an arbitrary subset of P . We show here that $S^\ell = \downarrow \max(S^\ell)$. It is straightforward by definition and independently from any assumption that $\downarrow \max(S^\ell) \subseteq S^\ell$. It remains to show that $S^\ell \subseteq \downarrow \max(S^\ell)$. Since (P, \leq) is *chain-complete*, then every $C \subseteq S^\ell$ has its join $\bigvee C \in P$. Hence, according to Lemma 2.13 and since $C \subseteq S^\ell$ then $\bigvee C \in S^\ell$. Thus every chain C in the sub-poset (S^ℓ, \leq) has an upper bound $\bigvee C \in S^\ell$. According to *Zorn's Lemma* (cf. equation 3), we have $S^\ell = \downarrow \max(S^\ell)$. Hence, (P, \leq) is a complete meet-multisemilattice.

In order to demonstrate the other statements of the theorem, one can follow the same steps to show that $S^u = \uparrow \min(S^u)$ using Zorn's Lemma on the dual poset of dually chain-complete posets. \square

Note 20. It is important to note that double chain-completeness is only a sufficient condition (under the *Axiom of Choice*) to have a complete multilattice but not a necessary one. Indeed, one can show that the poset depicted in Figure 8 is a complete multilattice (Remark that $\forall i \in \mathbb{N} : c_i \leq a_i, c_i \leq c_{i+1}, a_i \leq e_0$ and $a_i \leq e_1$) but not chain-complete since the chain $C = \{c_i \mid i \in \mathbb{N}\}$ has not a join. Indeed, $C^u = \{e_0, e_1\}$ which is an antichain (i.e. C^u has two minimal elements).

7. Pattern Multistructure

Definition 7.1 proposes a new structure that lies between *pattern setups* which rely only on arbitrary posets with no additional property and *pattern structures* which demand a *meet* for every subset of $\delta[\mathcal{G}]$.

Definition 7.1. A pattern setup $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \sqsubseteq), \delta)$ is said to be a **pattern multistructure** if every subset of $\delta[\mathcal{G}]$ has all its multi-infima, i.e. :

$$(\forall A \subseteq \mathcal{G}) \text{ cov}(A) = \downarrow \text{cov}^*(A)$$

A pattern multistructure adds an additional condition on a pattern setup which is the following: knowing maximal common descriptions covering all elements of a set of objects A allows us to deduce using the order \sqsubseteq every single covering description. Please note that using the notation of multi-infima, $\text{cov}^*(A)$ is given by $\text{minf}(\delta[A])$. It is clear that all pattern structures are by definition pattern multi-structures.

Graphs ordered by subgraph isomorphism relation introduced in Kuznetsov (1999) induce a pattern multistructure on the set of graphs, but not a pattern structure (a pattern structure is induced on sets of graphs incomparable wrt. subgraph isomorphism). Same remark holds for sequential patterns (Buzmakov et al. 2016; Codocedo et al. 2017). This is under the assumption of the existence of a largest element \top subsumed by all sequences/graphs (see Example 7.2).

Note 21. Note that in a pattern multistructure the empty set $\emptyset \subseteq \mathcal{D}$ has all its *multi-infima*. Since $\emptyset^\ell = \mathcal{D}$, the set \mathcal{D} has all its maximal elements (i.e. $\mathcal{D} = \downarrow \text{max}(\mathcal{D})$) or in other words every chain in $(\mathcal{D}, \sqsubseteq)$ is upper-bounded.

Example 7.2. Reconsider the pattern setup presented in Example 3.3. Since only finite sequences are considered in the description space, we have: $\text{cov}^*(\emptyset) = \emptyset$ even if $\text{cov}(\emptyset) = \mathcal{D}$. Thus, the considered pattern setup in Example 3.3 is *not a pattern multistructure* due to the empty set (recall that for a nonempty $A \subseteq \mathcal{G}$, $\delta[A]^\ell$ is finite and thus $\delta[A]$ has all its multi-infima). The common trick to handle the empty set is to enrich \mathcal{D} with an additional largest element $\top \triangleq \bigvee \mathcal{D}$ if it does not exist. In such a case, we have $\text{cov}^*(\emptyset) = \{\top\}$.

Let us now reconsider the question investigated at the end of section 5: “*What is the link between maximal covering descriptions and upper-approximations extents in a pattern multistructure*”. Before stating Theorem 7.4 answering this question, we shall state the following Lemma.

Lemma 7.3. Let (P, \leq) and (Q, \leq) be two posets and let $f : P \rightarrow Q$ be an order-reversing mapping. We have for any $S \subseteq P$ that $\uparrow f[\downarrow S] = \uparrow f[S]$.

Proof. Recall that \uparrow and \downarrow are closure operator (cf. Lemma 2.18). Let us start by showing that $\uparrow f[S] \subseteq \uparrow f[\downarrow S]$. Since $S \subseteq \downarrow S$, we have $f[S] \subseteq f[\downarrow S]$. Since \uparrow is monotone, we have $\uparrow f[S] \subseteq \uparrow f[\downarrow S]$. It remains to show that $\uparrow f[\downarrow S] \subseteq \uparrow f[S]$. Let $u \in \uparrow f[\downarrow S]$, that is $\exists v \in f[\downarrow S]$ s.t. $v \sqsubseteq u$. Since $v \in f[\downarrow S]$, then $\exists x \in \downarrow S$ s.t. $v = f(x)$. Hence $\exists y \in S$ s.t. $x \leq y$. Using the fact that f is an anti-embedding, we obtains that $f(y) \sqsubseteq f(x) \sqsubseteq u$. In other words, $\exists w \in f[S]$ s.t. $w \sqsubseteq u$. This is equivalent to say that $u \in \uparrow f[S]$. We conclude hence that $\uparrow f[\downarrow S] \subseteq \uparrow f[S]$. \square

Theorem 7.4. For any pattern multistructure \mathbb{P} we have:

$$(\forall A \subseteq \mathcal{G}) \quad \bar{A} = \min(\text{ext}[\text{cov}^*(A)])$$

Proof. The proof of the theorem is a straightforward application of Lemma 2.7 and Lemma 7.3. Let $A \subseteq \mathcal{G}$, since \mathbb{P} is a pattern multistructure, then:

$$\begin{aligned} \text{cov}(A) = \downarrow \max(\text{cov}(A)) &\Rightarrow \text{ext}[\text{cov}(A)] = \text{ext}[\downarrow \max(\text{cov}(A))] \\ &\Rightarrow \uparrow \text{ext}[\text{cov}(A)] = \uparrow \text{ext}[\downarrow \max(\text{cov}(A))] \end{aligned}$$

Since $\text{ext} : \mathcal{D} \rightarrow \wp(\mathcal{G})$ is an order reversing, then using Lemma 7.3 we have:

$$\begin{aligned} \uparrow \text{ext}[\downarrow \max(\text{cov}(A))] &= \uparrow \text{ext}[\max(\text{cov}(A))] && \Rightarrow \\ \uparrow \text{ext}[\text{cov}(A)] &= \uparrow \text{ext}[\max(\text{cov}(A))] && \Rightarrow \\ \min(\uparrow \text{ext}[\text{cov}(A)]) &= \min(\uparrow \text{ext}[\max(\text{cov}(A))]) \end{aligned}$$

Using Lemma 2.7 we obtain $\min(\text{ext}[\text{cov}(A)]) = \min(\text{ext}[\max(\text{cov}(A))])$. That is, $\bar{A} = \min(\text{ext}[\max(\text{cov}(A))])$. \square

Another important remark related to Example 5.8 is the fact that the support-closed patterns in a pattern setup does not hold all the information about the definable sets. Theorem 7.5 states that this is no longer the case for pattern multistructures.

Theorem 7.5. Given a pattern multistructure \mathbb{P} for which the set of support-closed patterns is \mathcal{D}^* (cf. Proposition 5.3), we have:

$$\mathbb{P}_{\text{ext}} = \text{ext}[\mathcal{D}^*]$$

Proof. Recall that $\mathcal{D}^* = \bigcup_{B \subseteq \mathcal{G}} \text{ext}[\text{cov}^*(B)]$. Since $\mathcal{D}^* \subseteq \mathcal{D}$ and by definition $\mathbb{P}_{\text{ext}} = \text{ext}[\mathcal{D}]$. It is clear that $\text{ext}[\mathcal{D}^*] \subseteq \text{ext}[\mathcal{D}]$. It remains to show that $\text{ext}[\mathcal{D}] \subseteq \text{ext}[\mathcal{D}^*]$. Let $A \in \text{ext}[\mathcal{D}]$, since \mathbb{P} is a pattern multistructure then $\text{cov}^*(A) = \downarrow \text{cov}(A)$. Let $d \in \text{cov}(A)$ s.t. $A = \text{ext}(d)$ (we have $A \in \text{ext}[\mathcal{D}]$). Since \mathbb{P} is a pattern multistructure then we have a support-closed pattern $d^* \in \text{cov}^*(A) \subseteq \mathcal{D}^*$ s.t. $d \sqsubseteq d^*$. Hence, $\text{ext}(d^*) \subseteq \text{ext}(d)$. Moreover, since $\text{cov}^*(A) \subseteq \text{cov}(A)$, we have $d^* \in \text{cov}(A)$. Therefore, $A = \text{ext}(d) \subseteq \text{ext}(d^*)$. We obtain thus $A = \text{ext}(d) = \text{ext}(d^*)$, that is $A \in \text{ext}[\mathcal{D}^*]$. \square

Similarly to Theorem 4.7 for pattern structures, Theorem 7.6 connects *multilattices* with *pattern multistructures*. It states that (complete) meet-multisemilattices are to pattern multistructures what (complete) lattices are to pattern structures.

Theorem 7.6. Let $\underline{\mathcal{D}} = (\mathcal{D}, \sqsubseteq)$ be a poset, the following properties are equivalent:

- For any finite set $\mathcal{G} \neq \emptyset$ and any $\delta \in \mathcal{D}^{\mathcal{G}}$, $(\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is a pattern multistructure.
- $\underline{\mathcal{D}}$ is a meet-multisemilattice having all its maximal elements (i.e. $\mathcal{D} = \downarrow \max(\mathcal{D})$)

The following properties are equivalent:

- For any set $\mathcal{G} \neq \emptyset$ and any $\delta \in \mathcal{D}^{\mathcal{G}}$, $(\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is a pattern multistructure.
- $\underline{\mathcal{D}}$ is a complete meet-multisemilattice.

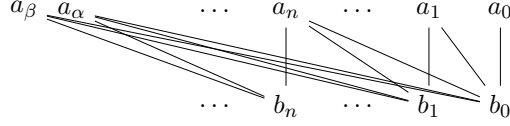


Figure 9. A complete multilattice with an infinite antichain (Proof of proposition 7.7)

Proof. Recall that $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is a pattern multistructure *iff* any subset $S \subseteq \delta[\mathcal{G}]$ has all its multi-infima. The proof of this theorem is almost the same as the one of Theorem 4.7 where the existence of the meet is replaced by the existence of all multi-infima. \square

Last but not least, we have seen in section 4 that in the case of a pattern structure, $(\mathbb{P}_{ext}, \sqsubseteq)$ is a complete lattice. One can say that the property of having the infimum in the description space has been transferred to the poset of definable sets thanks to extent operator. When it comes to a pattern setup on finite set of objects, it is clear that $(\mathbb{P}_{ext}, \sqsubseteq)$ is a complete multilattice since it is finite. However, does this property still hold for the case of infinite set of objects? Unfortunately, the answer is negative as stated in Proposition 7.7. This proposition tells also that not all definable sets above A in a pattern multistructure are above at least one upper-approximation of A .

Proposition 7.7. *There exists a pattern multistructure $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \sqsubseteq), \delta)$ such that $(\mathbb{P}_{ext}, \sqsubseteq)$ is not a join-multisemilattice in which: $(\exists A \subseteq \mathcal{G}) \uparrow \overline{A} \cap \mathbb{P}_{ext} \neq \uparrow A \cap \mathbb{P}_{ext}$.*

Proof. Consider the pattern setup $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \sqsubseteq), \delta)$ where $(\mathcal{D}, \sqsubseteq)$ is the complete multilattice depicted in Fig. 9. We have:

- $(\forall i, j \in \mathbb{N}) i \leq j \Leftrightarrow b_i \sqsubseteq a_j$.
- $(\forall i \in \mathbb{N}) b_i \sqsubseteq a_\alpha$ and $b_i \sqsubseteq a_\beta$.

Since $(\mathcal{D}, \sqsubseteq)$ is a complete multilattice (i.e. it is chain-finite), then \mathbb{P} is a pattern multistructure. Consider now an infinite set $\mathcal{G} = \{g_i \mid i \in \mathbb{N}\} \cup \{g_\alpha, g_\beta\}$. The mapping δ is given by: $\delta(g_\alpha) = a_\alpha$, $\delta(g_\beta) = a_\beta$ and $(\forall i \in \mathbb{N}) \delta(g_i) = a_i$. To show that the poset $(\mathbb{P}_{ext}, \sqsubseteq)$ is not a join-multisemilattice we need to consider two definable sets in \mathbb{P}_{ext} and show that the set of their common upper-bounds in \mathbb{P}_{ext} does not have all its minimal elements. Let us compute *ext* for every $d \in \mathcal{D}$:

- $ext(a_\alpha) = \{g_\alpha\}$ and $ext(a_\beta) = \{g_\beta\}$.
- $(\forall i \in \mathbb{N}) ext(a_i) = \{g_i\}$ and $(\forall i \in \mathbb{N}) ext(b_i) = \{g_\alpha, g_\beta\} \cup \{g_j \mid j \geq i\}$.

Consider now the set of definable sets $\{\{g_\alpha\}, \{g_\beta\}\}$, it is clear that the set of their common upper-bounds (in \mathbb{P}_{ext}) is given by:

$$\{\{g_\alpha\}, \{g_\beta\}\}^u = \{ \{g_\alpha, g_\beta\} \cup \{g_j \mid j \geq i\} \mid i \in \mathbb{N} \}$$

The set of upper bounds is hence an infinitely descending chain and hence does not have a minimal element, in other words: $min(\{\{g_\alpha\}, \{g_\beta\}\}^u) = \emptyset$. Hence, $(\mathbb{P}_{ext}, \sqsubseteq)$ is not a join-multisemilattice. The proof of the second part of the proposition is straightforward. Indeed, consider the non-definable set $A = \{g_\alpha, g_\beta\}$. We do have: $\uparrow A \cap \mathbb{P}_{ext} = ext[cov(A)] = \{ \{g_\alpha, g_\beta\} \cup \{g_j \mid j \geq i\} \mid i \in \mathbb{N} \}$. Hence, $\overline{A} = min(\uparrow A \cap \mathbb{P}_{ext}) = \emptyset$. That is $\uparrow A \cap \mathbb{P}_{ext} \neq \uparrow \overline{A} \cap \mathbb{P}_{ext}$. \square

8. Pattern Setup and Pattern Multistructure Completions

Example 3.3 presents a pattern setup which is not a pattern structure. However, in FCA and pattern structure literature, it is recurrent to talk about sequential pattern structures (Buzmakov et al. 2016; Codocedo et al. 2017). In fact, instead of sequences, sets of sequences are considered which induce a richer description space. Same trick has been even used in the first paper introducing pattern structures (Ganter and Kuznetsov 2001) concerning graph description space ordered by subgraph isomorphism. Such a technique that embeds a poset into another is called a *completion* (see Definition 2.16). Different natural completions exist in the literature. For instance, the Dedekind-MacNeille completion (Davey and Priestley 2002) takes an arbitrary poset to the smallest complete lattice containing it. The usual trick used in FCA and Pattern Structure literature to augment a base pattern setup to a pattern structure is tightly linked to the antichain completion presented below.

Definition 8.1. The **antichain completion** of (P, \leq) is the poset $(\mathcal{A}(P), \preceq)$ s.t.:

- $\mathcal{A}(P)$ is the set of all antichains of (P, \leq) .
- The order \preceq is given by $(\forall A, B \in \mathcal{A}(P)) A \preceq B \Leftrightarrow \downarrow A \subseteq \downarrow B$.⁵
- The order embedding φ from (P, \leq) to $(\mathcal{A}(P), \preceq)$ is given by

$$\varphi : P \rightarrow \mathcal{A}(P), a \mapsto \{a\}$$

Boldi and Vigna (2016) and Crampton and Loizou (2001) had discussed the properties of such a completion. In fact, when P have the **ACC**, $(\mathcal{A}(P), \preceq)$ is a *distributive lattice*, where the meet and the join are given by $S_1 \wedge S_2 = \max(\downarrow S_1 \cap \downarrow S_2)$ and $S_1 \vee S_2 = \max(S_1 \cup S_2)$, respectively. Moreover, $(\mathcal{A}(P), \preceq)$ is always a \vee -semilattice whatever the nature of the poset (P, \leq) , but not necessarily a lattice. Boldi and Vigna (2016) formulated a sufficient and necessary condition in order to have $(\mathcal{A}(P), \preceq)$ be a lattice: $\forall A, B \in \mathcal{A}(P) \exists C \in \mathcal{A}(P) \downarrow A \cap \downarrow B = \downarrow C$.

We take the opportunity here to underline an important link between the antichain completion and multilattices. Before expliciting this link in Proposition 8.3, let us take a close look to the following Lemma.

Lemma 8.2. *Let (P, \leq) be an arbitrary poset and let $\mathcal{A}(P)$ be the set of its antichain. We have $\forall S \subseteq P : (\exists C \in \mathcal{A}(P)) S = \downarrow C \Rightarrow C = \max(S)$.*

Proof. The case of $S = \emptyset$ is trivial since $\downarrow \emptyset = \emptyset$ and $\max(\emptyset) = \emptyset$. Let be a nonempty set $S \subseteq P$ s.t. $(\exists C \in \mathcal{A}(P)) S = \downarrow C$. Let us show that $C = \max(S)$:

- $C \subseteq \max(S)$: let $c \in C \subseteq S$, suppose that $c \notin \max(S)$ that is $\exists x \in S$ s.t. $c < x$. Since $S = \downarrow C$ then $\exists c_2 \in C$ s.t. $x \leq c_2$. Thus $\exists c_2 \in C$ such that $c > c_2$ which is a contradiction with the fact that C is an antichain.
- $C \supseteq \max(S)$: Suppose $\exists a \in \max(S)$ s.t. $a \notin C$. We have $a \in S = \downarrow C$, that is: $\exists c \in C$ s.t. $a < c$ (since $a \notin C$). However, since $S = \downarrow C$ then $C \subseteq S$. Thus, $\exists c \in S$ s.t. $a < c$ which is in contradiction with the fact that $a \in \max(S)$.

This concludes the proof. \square

⁵Note that \preceq does not induce an order in $\wp(P)$, but just a *pre-order*, since the *anti-symmetry* does not hold (see Crampton and Loizou (2001)). Indeed, consider poset $(\{a, b\}, \leq)$ where $a \leq b$. Since $\downarrow \{a, b\} = \downarrow \{b\} = \{a, b\}$, we have $\{a, b\} \preceq \{b\}$ and $\{b\} \preceq \{a, b\}$. Yet, $\{a, b\} \neq \{b\}$. Therefore, \preceq does not induce an antisymmetric relation on $(\wp(\{a, b\}), \preceq)$ but it is still reflexive and transitive (i.e. \preceq induce a preorder on $\wp(\{a, b\})$).

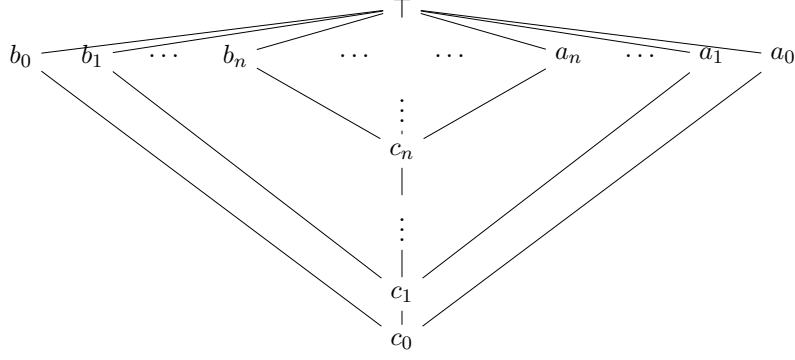


Figure 10. The Antichain completion of this complete lattice is not even a meet-semilattice

Proposition 8.3. *Let (P, \leq) be an arbitrary poset and let $(\mathcal{A}(P), \leq)$ be its antichain completion. If $(\mathcal{A}(P), \leq)$ is a lattice then (P, \leq) is a meet-multisemilattice. Moreover, if $(\mathcal{A}(P), \leq)$ has a top element then $P = \downarrow \max(P)$.*

Proof. Let us start by showing the first property that is if $(\mathcal{A}(P), \leq)$ is a lattice then (P, \leq) is meet-multisemilattice. We have $(\mathcal{A}(P), \leq)$ is a lattice. Then, following Boldi and Vigna (2016), we have:

$$\forall A \subseteq \mathcal{A}(P) \quad \exists C \in \mathcal{A}(P) \quad \downarrow A \cap \downarrow B = \downarrow C$$

More generally:

$$\forall S \subseteq \mathcal{A}(P) \text{ finite and nonempty} \quad \exists C \in \mathcal{A}(P) \quad \bigcap_{A \in S} \downarrow A = \downarrow C$$

Since C is an antichain, we have $C = \max(\bigcap_{A \in S} \downarrow A)$ (Lemma 8.2), that is:

$$\forall S \subseteq \mathcal{A}(P) \text{ finite and non empty} \quad \bigcap_{A \in S} \downarrow A = \downarrow \max\left(\bigcap_{A \in S} \downarrow A\right) \quad (4)$$

Let be $S \subseteq P$ be a non empty finite subset. We need to show that S has all its multi-infima. That is $S^\ell = \downarrow \max(S^\ell)$. We have $S^\ell = \bigcap_{s \in S} \downarrow \{s\}$. Since, $(\mathcal{A}(P), \leq)$ is a lattice we have according to equation (4):

$$S^\ell = \bigcap_{s \in S} \downarrow \{s\} = \downarrow \max(S^\ell)$$

For the second part of the proposition, consider that $\mathcal{A}(P)$ has a top elements. That is $\exists C \in \mathcal{A}(P) \quad P = \downarrow C$. Hence, we have $P = \downarrow \max(P)$ since C is an antichain (Lemma 8.2). In other words, \emptyset has all its multi-infima. \square

Note 22. One should note that the converse of Proposition 8.3. In fact, one can create complete lattices for which the antichain completion is not even a lattice. Fig. 10 depicts such a complete lattice. Indeed, for antichains $A = \{a_i \mid i \in \mathbb{N}\}$ and $B = \{b_i \mid i \in \mathbb{N}\}$, we have $\downarrow A \cap \downarrow B = \{c_i \mid i \in \mathbb{N}\}$. Hence, there is no antichain $D \in \mathcal{A}(P)$ s.t. $\{c_i \mid i \in \mathbb{N}\} = \downarrow D$ making the antichain completion not a meet-semilattice.

8.1. On Pattern Setups Antichain Completions

The main purpose of transforming a pattern setup to another one is to augment it to a pattern structure in order to use the different results related to this latter structure. We define below the most common trick used in the FCA literature which can be called *pattern setup antichain completion*.

Definition 8.4. Let $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ be a *pattern setup*, the **antichain completion** of \mathbb{P} is the *pattern setup* denoted by \mathbb{P}^∇ and given by:

$$\mathbb{P}^\nabla = (\mathcal{G}, (\mathcal{A}(\mathcal{D}), \leq), \delta^\nabla : g \mapsto \{\delta(g)\})$$

Earlier in this section, we have mentioned that $(\mathcal{A}(\mathcal{D}), \leq)$ is a lattice when $(\mathcal{D}, \sqsubseteq)$ is a finite poset (i.e. a sufficient condition). However, given an arbitrary pattern setup on an infinite description space, \mathbb{P}^∇ is not always guaranteed to be a pattern structure. Theorem 8.5 gives a necessary and sufficient condition on \mathbb{P} that makes \mathbb{P}^∇ a *pattern structure*. Here ext^∇ and int^∇ denote extent and intent of \mathbb{P}^∇ , respectively.

Theorem 8.5. *Let $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ be a pattern setup, the antichain completion of \mathbb{P} is a pattern structure if and only if \mathbb{P} is a pattern multistructure. Moreover:*

$$(\forall S \in \mathcal{A}(\mathcal{D})) \ ext^\nabla(S) = \bigcap \ ext[S] \quad (\forall A \subseteq \mathcal{G}) \ int^\nabla(A) = \min f(\delta[A]) = cov^*(A)$$

We have $\mathbb{P}_{ext}^\nabla = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$ and $\bigcap \emptyset = \mathcal{G} \in \mathbb{P}_{ext}^\nabla$.

Proof. Let us show that:

$$\mathbb{P} \text{ is a pattern multistructure} \iff \mathbb{P}^\nabla \text{ is a pattern structure}$$

Recall that \mathbb{P}^∇ is a pattern structure **iff** every subset of $\delta^\nabla[\mathcal{G}]$ has a meet in $(\mathcal{A}(\mathcal{D}), \leq)$. For $A \subseteq \mathcal{G}$ we have:

$$\delta^\nabla[A]^\ell = \{S \in \mathcal{A}(\mathcal{D}) \mid (\forall g \in A) S \subseteq \downarrow \delta(g)\} = \{S \in \mathcal{A}(\mathcal{D}) \mid S \subseteq \delta[A]^\ell\}$$

where $\delta[A]^\ell$ and $\delta^\nabla[A]^\ell$ denote respectively the lower bounds of $\delta[A]$ w.r.t. \sqsubseteq and the lower bounds of $\delta^\nabla[A]$ w.r.t. \leq (recall that $\delta[A]^\ell = \bigcap_{g \in A} \downarrow \delta(g)$). In this proof \downarrow refers to the down-closure related to \sqsubseteq .

We show each implication independently:

- (\Rightarrow) Let $A \subseteq \mathcal{G} : \delta[A]^\ell = \downarrow \max(\delta[A]^\ell)$. Thus $\delta^\nabla[A]^\ell = \{S \in \mathcal{A}(\mathcal{D}) \mid S \subseteq \downarrow \max(\delta[A]^\ell)\} = \{S \in \mathcal{A}(\mathcal{D}) \mid S \leq \max(\delta[A]^\ell)\}$. Since $\max(\delta[A]^\ell) \in \mathcal{A}(\mathcal{D})$, so $\max(\delta[A]^\ell)$ is the meet of $\delta^\nabla[A]$ in $\mathcal{A}(\mathcal{D})$.
- (\Leftarrow) \mathbb{P}^∇ is a pattern structure is **equivalent** to say: $\forall A \subseteq \mathcal{G}$, $\delta^\nabla[A]$ has a meet $M \in \mathcal{A}(\mathcal{D})$. That is, $\exists M \in \delta^\nabla[A]^\ell$ for $A \subseteq \mathcal{G} : \forall S \in \mathcal{A}(\mathcal{D}) : S \subseteq \delta[A]^\ell \Leftrightarrow S \subseteq \downarrow M$. Particularly, for $S = \{d\}$ with $d \in \mathcal{D}$, we deduce that $\forall d \in \delta[A]^\ell : d \in \downarrow M$. Thus, $\delta[A]^\ell \subseteq \downarrow M$. Moreover, since $M \subseteq \delta[A]^\ell$ ($M \in \delta^\nabla[A]^\ell$) and \downarrow is a closure operator on $(\wp(\mathcal{D}), \subseteq)$ we have by monotony $\downarrow M \subseteq \delta[A]^\ell \subseteq \downarrow M$ (note that $\downarrow \delta[A]^\ell = \delta[A]^\ell$). We conclude that we have $\delta[A]^\ell = \downarrow M$. Using Lemma 8.2 we obtain $\delta[A]^\ell = \downarrow \max(\delta[A]^\ell)$.

We conclude the equivalence.

Let us now determine int^∇ and ext^∇ . The previous proof has shown that for $A \subseteq \mathcal{G}$ the meet of $\delta^\nabla[A]$ is $max(\delta[A]^\ell)$. i.e.:

$$int^\nabla(A) = max(\delta[A]^\ell) = cov^*(A)$$

For ext^∇ operator, let $S \in \mathcal{A}(\mathcal{D})$. We have:

$$\begin{aligned} ext^\nabla(S) &= \{g \in \mathcal{G} \mid S \leq \sigma(g)\} = \{g \in \mathcal{G} \mid S \subseteq \downarrow \delta(g)\} \\ &= \{g \in \mathcal{G} \mid (\forall d \in S) d \subseteq \delta(g)\} = \bigcap_{d \in S} ext(d) = \bigcap ext[S] \end{aligned}$$

Let us show that $\mathbb{P}_{ext}^\nabla = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$. By definition of ext^∇ , the property $\mathbb{P}_{ext}^\nabla \subseteq \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$ holds. For the inverse inclusion, it is sufficient to show that $\mathbb{P}_{ext} \subseteq \mathbb{P}_{ext}^\nabla$ (since $(\mathbb{P}_{ext}^\nabla, \subseteq)$ is closed under intersection). Let $A \in \mathbb{P}_{ext}$. $\exists d \in \mathcal{D}$ s.t. $A = ext(d)$. Since $\{d\} \in \mathcal{A}(\mathcal{D})$, and $ext^\nabla(\{d\}) = ext(d) = A$. We conclude that $A \in \mathbb{P}_{ext}^\nabla$. Hence, $\mathbb{P}_{ext}^\nabla = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$. \square

8.2. On Pattern Setups Direct Completions

There is a completion that transforms any pattern setup to a pattern structure without demanding any additional property.

Theorem 8.6. *The **direct completion** of $\mathbb{P} = (\mathcal{G}, \underline{\mathcal{D}}, \delta)$ is the pattern structure:*

$$\mathbb{P}^\blacktriangleright = (\mathcal{G}, (\wp(\mathcal{D}), \subseteq), \delta^\blacktriangleright : g \mapsto \downarrow \delta(g))$$

Where $(\forall S \subseteq \mathcal{D}) ext^\blacktriangleright(S) = \bigcap ext[S]$ and $(\forall A \subseteq \mathcal{G}) int^\blacktriangleright(A) = cov(A) = \delta[A]^\ell$. The set of definable sets is given by $\mathbb{P}_{ext}^\blacktriangleright = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$.

Proof. Let us show that the pattern setup $\mathbb{P}^\blacktriangleright$ is a pattern structure. Let $A \subseteq \mathcal{G}$. We need to show that $\delta^\nabla[A]$ has a meet in $(\wp(\mathcal{D}), \subseteq)$. We have:

$$\delta^\nabla[A]^\ell = \{S \subseteq \mathcal{D} \mid (\forall g \in A) S \subseteq \downarrow \delta(g)\} = \{S \subseteq \mathcal{D} \mid S \subseteq \delta[A]^\ell\}$$

Since $\delta[A]^\ell \in \delta^\nabla[A]^\ell$, we conclude $\delta[A]^\ell$ is the meet of $\delta^\nabla[A]^\ell$. Hence:

$$int^\blacktriangleright(A) = \delta[A]^\ell = cov(A)$$

For the extent operator ext^\blacktriangleright , let $S \in \wp(\mathcal{D})$. We have

$$ext^\blacktriangleright(S) = \{g \in \mathcal{G} \mid S \subseteq \downarrow \delta(g)\} = \{g \in \mathcal{G} \mid (\forall d \in S) d \subseteq \delta(g)\} = \bigcap ext[S]$$

Let us show that $\mathbb{P}_{ext}^\blacktriangleright = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$. Thanks to the definition of ext^\blacktriangleright , property $\mathbb{P}_{ext}^\blacktriangleright \subseteq \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$ holds. For the inverse inclusion, it is sufficient to show that $\mathbb{P}_{ext} \subseteq \mathbb{P}_{ext}^\blacktriangleright$ (since $(\mathbb{P}_{ext}^\blacktriangleright, \subseteq)$ is closed under intersection). Let $A \in \mathbb{P}_{ext}$, $\exists d \in \mathcal{D}$ s.t. $A = ext(d)$. We have $ext^\blacktriangleright(\{d\}) = \{g \in \mathcal{G} \mid \{d\} \subseteq \downarrow \delta(g)\} = \{g \in \mathcal{G} \mid d \subseteq \delta(g)\} = ext(d) = A$. We conclude that $A \in \mathbb{P}_{ext}^\blacktriangleright$ and $\mathbb{P}_{ext}^\blacktriangleright = \{\bigcap S \mid S \subseteq \mathbb{P}_{ext}\}$. \square

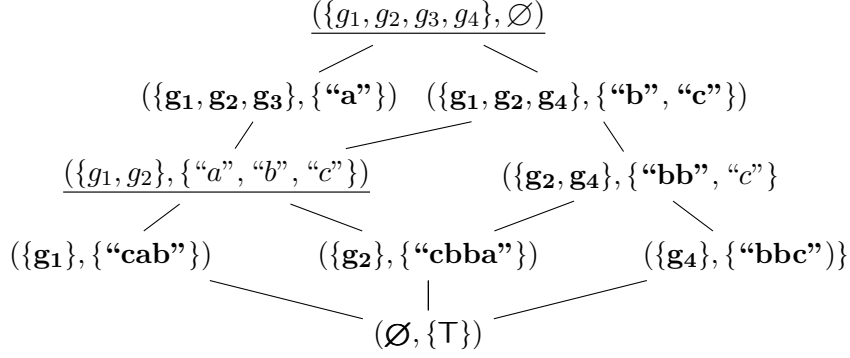


Figure 11. Concept lattice $\mathfrak{B}(\mathbb{P}^\nabla)$.

Example 8.7. Fig. 11 depicts the concept lattice $\mathfrak{B}(\mathbb{P}^\nabla)$ of the *antichain completion* of the pattern multistructure \mathbb{P} considered in Fig. 4 (i.e., the description space is augmented with the top element \top). For any concept (A, B) , descriptions $d \in B$ in **bold** are those which whose $ext(d) = A$. Please notice that while description “c” has for extent $\{g_1, g_2, g_4\}$, description “c” does belong to the concept related to the extent $\{g_2, g_4\}$. Another important remark, are the underlined concepts. They represent concepts that are related to the non definable sets in \mathbb{P} but still definable in \mathbb{P}^∇ , i.e. $\{g_1, g_2\}$ and $\{g_1, g_2, g_3, g_4\}$ in $\mathbb{P}_{ext}^\nabla \setminus \mathbb{P}_{ext}$. For instance, consider the intent of $\{g_1, g_2\}$ in the completion, each pattern d has extent $ext(d) \supseteq \{g_1, g_2\}$. Extent $\{g_1, g_2, g_3, g_4\}$ is *non-coverable* in \mathbb{P} and thus $int^\nabla(\{g_1, g_2, g_3, g_4\}) = max(cov(\{g_1, g_2, g_3, g_4\})) = max(\emptyset) = \emptyset$.

Note that while in Example 8.7, the size difference between the set of definable sets in the base pattern setup \mathbb{P}_{ext} and the set of definable sets in the antichain completion \mathbb{P}_{ext}^∇ is not large (i.e. $|\mathbb{P}_{ext}^\nabla \setminus \mathbb{P}_{ext}| = 2$). In some cases, the size of \mathbb{P}_{ext}^∇ can be exponentially larger than \mathbb{P}_{ext} . Consider, for instance, the following example:

Example 8.8. Let $n \in \mathbb{N}$ with $n \geq 3$. We denote by $[n]$ the subset $[n] = \{1, 2, \dots, n\}$. Let $\mathbb{P} = (\mathcal{G}, (\mathcal{D}, \subseteq), \delta)$ be the pattern setup with $\mathcal{G} = \{g_i\}_{i \in [n]}$,

$$\mathcal{D} = \{\{i\} \mid i \in [n]\} \cup \{[n] \setminus \{i\} \mid i \in [n]\}$$

and the mapping $\delta : g_i \mapsto [n] \setminus \{i\}$ for all $i \in [n]$. One can verify that we have $\mathbb{P}_{ext} = \{\{g_i\} \mid i \in [n]\} \cup \{\mathcal{G} \setminus \{g_i\} \mid i \in [n]\}$. Indeed, we have:

- $(\forall i \in [n]) \ ext([n] \setminus \{i\}) = \{g_j \mid [n] \setminus \{j\} \subseteq \delta(g_j)\} = \{g_j \mid [n] \setminus \{j\} \subseteq [n] \setminus \{i\}\} = \{g_i\}$.
- $(\forall i \in [n]) \ ext(\{i\}) = \{g_j \mid \{i\} \subseteq \delta(g_j)\} = \{g_j \mid \{i\} \subseteq [n] \setminus \{j\}\} = \mathcal{G} \setminus \{g_i\}$.

Hence, we have $|\mathbb{P}_{ext}| = 2n$. However, according to Theorem 8.5, we have

$$\mathbb{P}_{ext}^\nabla = \left\{ \bigcap S \mid S \subseteq \mathbb{P}_{ext} \right\} = \wp(\mathcal{G})$$

since \mathbb{P}_{ext} contains all the coatoms of $\wp(\mathcal{G})$ (i.e. $(\forall g \in \mathcal{G}) \ \mathcal{G} \setminus \{g\} \in \mathbb{P}_{ext}$) and the powerset lattice is coatomistic. It follows that $|\mathbb{P}_{ext}^\nabla| = 2^n$. In other words, \mathbb{P}_{ext}^∇ is exponentially larger than \mathbb{P}_{ext} . One should notice that the new description space associated to \mathbb{P}^∇ is (order-)isomorphic to $(\wp([n]), \subseteq)$.

9. Conclusion and Discussion

In this paper, we provided a better understanding of the pattern setup framework. We have shown that while pattern structures demand a strong condition on the partially ordered set of descriptions, pattern setups do not require any additional property on the description space, which makes them rather permissive. We have introduced a new framework, namely pattern multistructure, lying between both structures. Informally, pattern multistructures demands that the set of maximal common description resume properly the set of common descriptions of any subset of objects. Analogously to pattern structures, pattern multistructures are tightly linked to multilattices. We have shown also that the usual *antichain completion* used in FCA literature to transform pattern setups, like sequence of itemsets ones, to pattern structure is applicable if and only if the considered pattern setup is a pattern multistructure.

An important open problem we are thoroughly working on is the following: “Given an arbitrary pattern setup \mathbb{P} with a finite set of objects, generate its definable sets exhaustively and irredundantly”. If the pattern setup is a pattern structure, the literature abounds of algorithms solving such a problem (e.g. (Ganter 1984; Kuznetsov 1993; Outrata and Vychodil 2012)). However, no algorithm exists to solve such a problem for an arbitrary pattern setup. Indeed, the usual solution in FCA is to transform the pattern setup to a pattern structure via a completion (e.g. antichain completion for (Kuznetsov and Samokhin 2005; Buzmakov et al. 2016; Codocedo et al. 2017)). Such a solution could create a (exponentially) larger search space. Other algorithms in the literature tackles this problem by enumerating support-closed patterns (e.g. Yan, Han, and Afshar (2003)). These latter algorithms may generate twice the same definable sets, since two support-closed patterns could have the same extent. Solving this problem will be the subject of future research.

References

- Agrawal, Rakesh, and Ramakrishnan Srikant. 1995. “Mining sequential patterns.” In *Data Engineering*, 3–14. IEEE.
- Baixeries, Jaume, Mehdi Kaytoue, and Amedeo Napoli. 2012. “Computing Functional Dependencies with Pattern Structures.” In *CLA*, Vol. 972, 175–186.
- Baixeries, Jaume, Mehdi Kaytoue, and Amedeo Napoli. 2014. “Characterizing functional dependencies in formal concept analysis with pattern structures.” *Ann. Math. Artif. Intell.* 72 (1-2): 129–149.
- Belfodil, Aimene, Sergei O. Kuznetsov, and Mehdi Kaytoue. 2018. “Pattern Setups and Their Completions.” In *CLA*, Vol. 2123, 243–253.
- Belfodil, Aimene, Sergei O. Kuznetsov, Céline Robardet, and Mehdi Kaytoue. 2017. “Mining Convex Polygon Patterns with Formal Concept Analysis.” In *IJCAI*, 1425–1432.
- Benado, Mihail. 1955. “Les ensembles partiellement ordonnés et le théorème de raffinement de Schreier I Théorie des multistructures.” *Czechoslovak Math. J.* 5 (3): 308–344.
- Boldi, Paolo, and Sebastiano Vigna. 2016. “On the lattice of antichains of finite intervals.” *CoRR* abs/1510.03675. <https://arxiv.org/abs/1510.03675>.
- Boley, Mario, Tamás Horváth, Axel Poigné, and Stefan Wrobel. 2010. “Listing closed sets of strongly accessible set systems with applications to data mining.” *Theor. Comput. Sci.* 411 (3): 691–700.
- Buzmakov, Aleksey, Elias Egho, Nicolas Jay, Sergei O. Kuznetsov, Amedeo Napoli, and Chedy Raïssi. 2016. “On mining complex sequential data by means of FCA and pattern structures.” *Int. J. General Systems* 45 (2): 135–159.

- Buzmakov, Aleksey, Sergei O. Kuznetsov, and Amedeo Napoli. 2015. "Revisiting Pattern Structure Projections." In *ICFCA*, Vol. 9113 of *Lecture Notes in Computer Science*, 200–215.
- Codocedo, Víctor, Guillaume Bosc, Mehdi Kaytoue, Jean-François Boulicaut, and Amedeo Napoli. 2017. "A Proposition for Sequence Mining Using Pattern Structures." In *ICFCA*, .
- Codocedo, Víctor, and Amedeo Napoli. 2014. "Lattice-based biclustering using Partition Pattern Structures." In *ECAI 2014*, Vol. 263, 213–218.
- Cordero, Pablo, Gloria Gutiérrez, Javier Martínez, and Inmaculada Perez de Guzmán. 2004. "A new algebraic tool for automatic theorem provers." *Annals of Mathematics and Artificial Intelligence* 42 (4): 369–398.
- Crampton, Jason, and George Loizou. 2001. "The completion of a poset in a lattice of antichains." *International Mathematical Journal* 1 (3): 223–238.
- Davey, Brian A, and Hilary A Priestley. 2002. *Introduction to lattices and order*. Cambridge university press.
- Ferré, Sébastien, and Olivier Ridoux. 2000. "A Logical Generalization of Formal Concept Analysis." In *ICCS*, Vol. 1867 of *Lecture Notes in Computer Science*, 371–384.
- Ganter, Bernhard. 1984. "Two basic algorithms in concept analysis." *Technical report, Technische Hochschule Darmstadt* .
- Ganter, Bernhard, and Sergei O. Kuznetsov. 2001. "Pattern Structures and Their Projections." In *ICCS*, Vol. 2120 of *LNCS*, 129–142. Springer.
- Ganter, Bernhard, and Rudolf Wille. 1989. "Conceptual scaling." In *Applications of combinatorics and graph theory to the biological and social sciences*, 139–167.
- Ganter, Bernhard, and Rudolf Wille. 1999. *Formal Concept Analysis*. Springer.
- Garriga, Gemma C., Petra Kralj, and Nada Lavrac. 2008. "Closed Sets for Labeled Data." *Journal of Machine Learning Research* 9: 559–580.
- Kaytoue, Mehdi, Sergei O. Kuznetsov, and Amedeo Napoli. 2011. "Revisiting Numerical Pattern Mining with Formal Concept Analysis." In *IJCAI*, 1342–1347.
- Kuznetsov, Sergei O. 1993. "A Fast Algorithm for Computing All Intersections of Objects in a Finite Semi-lattice." *Nauchno-Tekhnicheskaya Informatsiya ser. 2* (1): 17–20.
- Kuznetsov, Sergei O. 1999. "Learning of Simple Conceptual Graphs from Positive and Negative Examples." In *PKDD*, 384–391.
- Kuznetsov, Sergei O. 2009. "Pattern Structures for Analyzing Complex Data." In *RSFDGrC*, 33–44.
- Kuznetsov, Sergei O., and Mikhail V. Samokhin. 2005. "Learning Closed Sets of Labeled Graphs for Chemical Applications." In *ILP*, Vol. 3625 of *LNCS*, 190–208.
- Lumpe, Lars, and Stefan E. Schmidt. 2015. "Pattern Structures and Their Morphisms." In *CLA*, 171–179.
- Martínez, J, G Gutiérrez, IP de Guzmán, and P Cordero. 2001. "Multilattices via multisemi-lattices." *Topics in applied and theoretical mathematics and computer science* 238–248.
- Martínez, Javier, Gloria Gutiérrez, Inmaculada Perez de Guzmán, and Pablo Cordero. 2005. "Generalizations of lattices via non-deterministic operators." *Discrete Mathematics* 295 (1-3): 107–141.
- Outrata, Jan, and Vilém Vychodil. 2012. "Fast algorithm for computing fixpoints of Galois connections induced by object-attribute relational data." *Inf. Sci.* 185 (1): 114–127.
- Pawlak, Zdzislaw. 1982. "Rough sets." *International Journal of Parallel Programming* 11 (5): 341–356.
- Roman, S. 2008. *Lattices and Ordered Sets*. Springer New York.
- Wang, Jianyong, and Jiawei Han. 2004. "BIDE: Efficient Mining of Frequent Closed Sequences." In *ICDE*, 79–90. IEEE Computer Society.
- Wille, Rudolf. 1982. "Restructuring lattice theory: an approach based on hierarchies of concepts." In *Ordered sets*, 445–470.
- Yan, Xifeng, Jiawei Han, and Ramin Afshar. 2003. "CloSpan: Mining Closed Sequential Patterns in Large Datasets." In *SDM*, 166–177. SIAM.
- Zaki, Mohammed Javeed. 2001. "SPADE: An Efficient Algorithm for Mining Frequent Sequences." *Machine Learning* 42 (1/2): 31–60.