Succinct System of Minimal Generators: A thorough study, limitations and new definitions

T. Hamrouni[1], S. Ben Yahia[1] and E. Mephu Nguifo[2]

[1] Computer Science department, Faculty of Sciences of Tunis, Tunis, Tunisia.
{*tarek.hamrouni, sadok.benyahia*}*@fst.rnu.tn*
[2] CRIL-CNRS, IUT de Lens, Lens, France.
mephu@cril.univ-artois.fr

Abstract. Minimal generators (MGs) are the smallest ones (*w.r.t.* the number of items) among equivalent itemsets sharing a common set of objects, while their associated closed itemset (CI) is the largest one. The pairs - composed by MGs and their associated CI - divide the itemset lattice into distinct equivalence classes. Such pairs were at the origin of various works related to generic association rule bases, concise representations, arbitrary boolean expressions, etc. Furthermore, the MG set presents some important properties like the order ideal. The latter helped some level-wise bottom-up and even slightly modified depth-first algorithms to efficiently extract interesting knowledge. Nevertheless, the inherent absence of a unique MG associated to a given CI motivates an in-depth study of the possibility of discovering a kind of redundancy within the MG set. This study was started by Dong *et al.* who introduced the succinct system of minimal generators (SSMG) as an attempt to eliminate the redundancy within this set. In this paper, we give a thorough study of the SSMG as formerly defined by Dong *et al*. Then, we show that the latter suffers from some drawbacks. After that, we introduce new definitions allowing to overcome the limitations of their work. Finally, an experimental evaluation shows that the SSMG makes it possible to eliminate without information loss an important number of *redundant* MGs.

1 Introduction

One efficient way to characterize the itemset lattice is to divide it into different equivalence classes [1]. The smallest elements (*w.r.t.* the number of items) in each equivalence class are called minimal generators (MGs) [2] (also referred to as **0**-free itemsets [3] and key patterns [4]) and the largest element is called a closed itemset (CI) [5]. The set of *frequent* CIs is among the first concise representations of the whole set of *frequent* itemsets that were introduced in the literature. This set have been extensively studied and tens of algorithms were proposed to efficiently extract it [4–11] (1). In the contrary, and despite the important role played by the MGs, they have been paid little attention. Indeed, the MG set is, in general, extracted as a means to achieve *frequent* itemset computations [1, 13], *frequent* CI computations [4, 6, 7], Iceberg lattice construction [14], etc. The use of the MGs was mainly motivated by their small sizes (they are hence the first elements to be reached in each equivalence class) and by the fact that the MG set

[1] A critical survey on *frequent* CI based algorithms can be found in [12].

verifies the order ideal property which clearly increased the efficiency of both level-wise bottom-up algorithms [4, 6, 14] and even slightly modified depth-first ones [7]. Nevertheless, some work has been done on the semantic advantages offered by the use of MGs. These works are mainly related to generic association rule bases [2, 15–17], concise representations [7, 18], arbitrary boolean expressions [19], etc.

Nevertheless, the inherent absence of a unique MG associated to a given CI motivates an in-depth study to try to discover a kind of redundancy within the MGs associated to a given CI. This study was started thanks to Dong *et al.* who recently note that some MGs associated to a given CI can be derived from other ones [20]. Indeed, they consider the set of the MGs by distinguishing two distinct categories: *succinct* MGs and *redundant* ones. Thus, Dong *et al.* introduce the succinct system of minimal generators (SSMG) as a concise representation of the MG set. They state that *redundant* MGs can be pruned out from the MG set since they can straightforwardly be inferred, without loss of information, using the information gleaned from *succinct* ones [20].

In this paper, we give a thorough study of the SSMG as formerly defined by Dong *et al.* [20]. Then, we show that the *succinct* MGs, as defined in [20], proves *not* to be an *exact* representation of the MG set (no loss of information *w.r.t.* *redundant* MGs) in contrary to authors' claims. Furthermore, we also show that the different SSMGs associated to an extraction context do not necessarily share the same size, in contrary to what was stated in [20]. After that, we introduce new definitions allowing to overcome the limitations of the work of Dong *et al*. Indeed, our definitions allow, on the one hand, the SSMG to be an *exact* representation and, on the other hand, the different SSMGs associated to an extraction context to have the same size. Finally, carried out experiments show that the SSMG makes it possible to eliminate without loss of information an important number of *redundant* MGs and, hence, to almost reach the ideal case: *only one succinct* MG per equivalence class.

The organization of the paper is as follows: Section 2 recalls some preliminary notions that will be used in the remainder of the paper. Section 3 presents a detailed formal study of the SSMG as formerly defined by Dong *et al.* [20], sketches its limitations, and gives new definitions allowing to go beyond the drawbacks of their work. Section 4 is dedicated to the related work. In Section 5, several experiments illustrate the utility of the SSMG towards eliminating redundancy within the MG set. Finally, Section 6 concludes this paper and points out our future work.

2 Preliminary notions

In this section, we present some notions that will be used in the following.

Definition 1. (EXTRACTION CONTEXT) *An extraction context is a triplet* $\mathcal{K} = (\mathcal{O}, \mathcal{I}, \mathcal{R})$, *where* $\mathcal{O}$ *represents a finite set of objects,* $\mathcal{I}$ *is a finite set of items and* $\mathcal{R}$ *is a binary (incidence) relation (i.e.,* $\mathcal{R} \subseteq \mathcal{O} \times \mathcal{I}$*). Each couple* $(o, i) \in \mathcal{R}$ *expresses that the object* $o \in \mathcal{O}$ *contains the item* $i \in \mathcal{I}$.

The closure operator $('')$ denotes the closure operator $\phi \circ \psi$ *s.t.* (ϕ, ψ) represents a couple of operators defined by $\psi : \mathcal{P}(\mathcal{I}) \rightarrow \mathcal{P}(\mathcal{O})$ *s.t.* $\psi(I) = \{o \in \mathcal{O} \mid \forall\ i \in I, (o, i) \in \mathcal{R}\}$ and $\phi : \mathcal{P}(\mathcal{O}) \rightarrow \mathcal{P}(\mathcal{I})$ *s.t.* $\phi(O) = \{i \in \mathcal{I} \mid \forall\ o \in O, (o, i) \in \mathcal{R}\}$ [21].

It induces an equivalence relation on the power set of items portioning it into distinct subsets called *equivalence classes* [1]. The largest element (*w.r.t.* the number of items) in each equivalence class is called a *closed itemset* (CI) [5] and the smallest ones are called *minimal generators* (MGs) [2]. The notions of *closed itemset* and of *minimal generator* are defined as follows:

Definition 2. (CLOSED ITEMSET) *An itemset $I \subseteq \mathcal{I}$ is said to be closed if and only if I'' = I [5]. The support of I, denoted by Supp(I), is equal to the number of objects in $\mathcal{K}$ that contain I. I is said to be frequent if Supp(I) is greater than or equal to a minimum support threshold, denoted minsupp.*

Definition 3. (MINIMAL GENERATOR) *An itemset $g \subseteq \mathcal{I}$ is said to be a minimal generator (*MG*) of a* CI *f, if and only if g'' = f and $\nexists\, g_1 \subset g$ s.t. g_1'' = f [2]. Thus, the set* MG$_f$ *of the* MGs *associated to a* CI *f is:* MG$_f$ = $\{g \subseteq \mathcal{I} \mid g'' = f \wedge \nexists\, g_1 \subset g$ *s.t.* g_1'' $= f\}$.

3 Succinct System of Minimal Generators

In this section, and as a first step, we study the main structural properties of the succinct system of minimal generators (SSMG) as formerly defined by Dong *et al* [20]. As a second step, we give some drawbacks of their work. Finally, we propose new definitions allowing to overcome these limitations.

3.1 A thorough study

Recently, Dong *et al.* note the existence of a certain form of redundancy within the set of the minimal generators (MGs) associated to a given closed itemset (CI), *i.e.*, that one can derive some MGs from the others. They, hence, presented a study [20] in which they split the set of MGs associated to a given CI into three distinct subsets. The formalization of these subsets, introduced in Definition 5, requires that we adopt a total order relation among itemsets defined as follows.

Definition 4. (TOTAL ORDER RELATION) *Let $\preceq$ be a total order relation among item literals, i.e., $\forall\, i_1,\, i_2 \in \mathcal{I}$, we have $i_1 \preceq i_2$ or $i_2 \preceq i_1$. This relation is extended to also cope with itemsets of different sizes by first considering their cardinality. This is done as follows: Let X and Y be two itemsets and i an item s.t. $i \notin X$ and $i \notin Y$. Let Card(X) and Card(Y) be the respective cardinalities of X and Y. We then have:*

- *Card(X) < Card(Y) $\Longrightarrow X \prec Y$.*
- *$X \preceq Y \Longleftrightarrow X \cup \{i\} \preceq Y \cup \{i\}$.*

Example 1. If we consider the lexicographic order as the total order relation $\preceq$, then [2]:

- $|d| < |be| \Longrightarrow d \prec be$.
- *abd* $\preceq$ *abe* $\Longleftrightarrow$ *abd* $\cup \{c\} \preceq$ *abe* $\cup \{c\}$ (*i.e.*, *abcd* $\preceq$ *abce*).

[2] We use a separator-free form for the sets, *e.g.*, *be* stands for $\{b, e\}$.

Definition 5. (MINIMAL GENERATORS' CLASSES) *The set* MG_f*, of the* MG*s associated to a* CI f*, can be divided into three distinct subsets as follows:*

1. $\mathrm{MG}\mathit{rep}_f = \{g \in \mathrm{MG}_f \mid \nexists\, g_1 \in \mathrm{MG}_f \text{ s.t. } g_1 \prec g\}$*: the subset* $\mathrm{MG}\mathit{rep}_f$ *contains the* ***representative*** MG *of* f*. This* MG *is the smallest one being given a total order relation* $\preceq$*.*
2. $\mathrm{MG}\mathit{can}_f = \{g \in \mathrm{MG}_f \mid g \notin \mathrm{MG}\mathit{rep}_f \wedge \forall\, g_1 \subset g, g_1 \in \mathrm{MG}\mathit{rep}_{f_1} \text{ with } f_1 = g_1''\}$*: the subset* $\mathrm{MG}\mathit{can}_f$ *contains the* ***canonical*** MG*s of* f*. A canonical* MG *is not the smallest one in* MG_f *and, hence, is not the representative* MG *of* f*. Nevertheless, all its subsets are the representative* MG*s of their respective closed itemsets, which are necessarily covered by* f*.*
3. $\mathrm{MG}\mathit{red}_f = \{g \in \mathrm{MG}_f \mid \exists\, g_1 \subset g, g_1 \notin \mathrm{MG}\mathit{rep}_{f_1} \text{ with } f_1 = g_1''\}$*: the subset* $\mathrm{MG}\mathit{red}_f$ *contains the* ***redundant*** MG*s of* f*. A redundant* MG *has at least one of its subsets which is not a representative* MG*.*

If a MG g *is a representative or a canonical one, then* g *is called a* ***succinct*** MG*. Hence, the set* $\mathrm{MG}\mathit{suc}_f$*, composed by the succinct* MG*s associated to the* CI f*, is equal to the union of* $\mathrm{MG}\mathit{rep}_f$ *and* $\mathrm{MG}\mathit{can}_f$*:* $\mathrm{MG}\mathit{suc}_f = \mathrm{MG}\mathit{rep}_f \cup \mathrm{MG}\mathit{can}_f$*. Hence,* $\mathrm{MG}\mathit{red}_f = \mathrm{MG}_f \backslash \mathrm{MG}\mathit{suc}_f$*.*

Example 2. Let us consider the extraction context $\mathcal{K}$ depicted by Figure 1 (Left). The total order relation $\preceq$ is set to the lexicographic order. Figure 1 (Right) shows, for each CI, the following information: its MGs, its *succinct* MGs and its support. In the fourth column, the *representative* MG is marked with bold letters. The others are hence *canonical* ones. Note that for $\mathbf{11}$ CIs, there are $\mathbf{23}$ MGs, from which only $\mathbf{13}$ are *succinct* ones ($\mathbf{11}$ are *representative* MGs and only $\mathbf{2}$ are *canonical* ones). The MG "ad" is a *representative* one, since it is the smallest MG, *w.r.t.* $\preceq$, among those of "$abcde$". Indeed, $ad \preceq ae$, $ad \preceq bd$ and $ad \preceq be$. The MG "e" is not the *representative* of its CI "cde", since $d \preceq e$. Nevertheless, its unique subset (*i.e.*, "$\emptyset$") is the *representative* MG of its CI "c". Hence, "e" is a *canonical* MG. Finally, the MG "bdg" is a *redundant* one, since at least one of its subsets is not a *representative* MG ("bg", for example).

The definition of the SSMG is as follows [20]:

Definition 6. (SUCCINCT SYSTEM OF MINIMAL GENERATORS) *Being given a total order relation* $\preceq$*, a succinct system of minimal generators (*SSMG*) consists of, for each* CI f*, the set* $\mathrm{MG}\mathit{rep}_f$ *containing its representative* MG *and, if there is any, the set* $\mathrm{MG}\mathit{can}_f$ *containing its canonical* MG*s.*

It is important to mention that, for a given extraction context, the SSMG is not unique since it closely depends on the choice of the total order relation $\preceq$ (*e.g.*, the lexicographic order, the ascending/descending support order, etc.).

In the remainder, the set of *representative* (resp. *canonical*, *redundant*, *succinct* and all) MGs extracted from a context $\mathcal{K}$ will be denoted $\mathcal{MG}\mathrm{rep}_{\mathcal{K}}$ (resp. $\mathcal{MG}\mathrm{can}_{\mathcal{K}}$, $\mathcal{MG}\mathrm{red}_{\mathcal{K}}$, $\mathcal{MG}\mathrm{suc}_{\mathcal{K}}$ and $\mathcal{MG}_{\mathcal{K}}$). The set of CIs extracted from $\mathcal{K}$ will be denoted $\mathcal{CI}_{\mathcal{K}}$. The letter $\mathcal{F}$ will be added to each denotation if the respective set is restricted to its *frequent* elements.

	a	b	c	d	e	f	g
1			×	×	×	×	×
2	×	×	×	×	×		
3	×	×	×			×	×
4	×	×	×	×	×		×

	CI	**MGs**	**Succinct MGs**	**Support**
1	*c*	∅	∅	4
2	*abc*	*a, b*	***a**, b*	3
3	*cde*	*d, e*	***d**, e*	3
4	*cg*	*g*	***g***	3
5	*cfg*	*f*	***f***	2
6	*abcde*	*ad, ae, bd, be*	***ad***	2
7	*abcg*	*ag, bg*	***ag***	2
8	*abcfg*	*af, bf*	***af***	1
9	*cdeg*	*dg, eg*	***dg***	2
10	*cdefg*	*df, ef*	***df***	1
11	*abcdeg*	*adg, aeg, bdg, beg*	***adg***	1

Fig. 1. (**Left**) An extraction context $\mathcal{K}$. (**Right**) The CIs extracted from $\mathcal{K}$ and for each one, the corresponding MGs, *succinct* MGs and support.

Proposition 1. *The total order relation* $\preceq$ *ensures the uniqueness of the representative* MG *associated to a given* CI. *Hence, the cardinality of the set of representative* MG*s is equal to that of* CI*s (i.e., Card(*$\mathcal{MG}\texttt{rep}_{\mathcal{K}}$*) = Card(*$\mathcal{CI}_{\mathcal{K}}$*)).*

The proof is trivial. Indeed, there is only one representative MG per equivalence class (see Definition 5).

Remark 1. The respective sizes of both sets $\mathcal{MG}\texttt{can}_{\mathcal{K}}$ and $\mathcal{MG}\texttt{red}_{\mathcal{K}}$ is closely related to the nature of the extraction context, *i.e.*, whether the objects are highly/weakly correlated. Nevertheless, if the set $\mathcal{MG}\texttt{can}_{\mathcal{K}}$ is empty, then the set $\mathcal{MG}\texttt{red}_{\mathcal{K}}$ is also empty (the reverse is not always true).

To show that the set $\mathcal{MG}\texttt{suc}_{\mathcal{K}}$ is an order ideal, we have to prove that all subsets of a *representative* MG are also *representative* ones. This is done thanks to Proposition 2 whose the proof requires Lemma 1.

Lemma 1. *[21] Let* X *and* Y *be two itemsets. If* $X'' = Y''$*, then* $\forall\ Z \subseteq \mathcal{I}$*,* $(X \cup Z)'' = (Y \cup Z)''$*.*

Proposition 2. *All subsets of a representative* MG *are also representative ones.*

Proof.
Let g *be a representative* MG *and* f *its closure. Suppose, we have* $g_1 \subset g$ *and* $g_1 \notin \text{MG}\texttt{rep}_{f_1}$ *with* $f_1 = g_1''$*. Let* g_2 *be the representative* MG *of* f_1*. Consequently,* $g_2 \prec g_1$*. Since,* $g_1'' = g_2''$*, then, according to Lemma 1, we have* $(g_1 \cup (g \setminus g_1))'' = (g_2 \cup (g \setminus g_1))''$ *and, hence,* $g'' = (g_2 \cup (g \setminus g_1))''$*. Let* g_3 *be equal to* $(g_2 \cup (g \setminus g_1))$*. According to the second case in Definition 4, we have* $g_3 \prec g$ *since* $g_2 \prec g_1$*. If* g_3 *is a* MG*, then* g *can not be a representative* MG *what is in contradiction with the initial assumption that* g *is a representative* MG*. If* g_3 *is not a* MG*, then it exists a* MG g_4 *such that* $g_4 \subset g_3$ *and* $g_4'' = g_3''$*. Since Card(*g_4*)* $<$ *Card(*g_3*), then* $g_4 \prec g_3$ *(according to the first case in Definition 4) and, hence,* $g_4 \prec g$*. This result is also in contradiction with the starting*

assumption. Thus, we can conclude that each subset of g is necessarily a representative MG.

Hence, according to Proposition 2, if f is a CI, then $\text{MG}\texttt{suc}_f = \text{MG}\texttt{rep}_f \cup \text{MG}\texttt{can}_f$ = $\{g \in \text{MG}_f \mid \forall\, g_1 \subset g,\, g_1 \in \text{MG}\texttt{rep}_{f_1} \text{ with } f_1 = g_1''\}$.

Thanks to Proposition 3, given below with its proof, we show that the *succinctness* of MGs is an anti-monotone constraint. Hence, the set $\mathcal{MG}\texttt{suc}_{\mathcal{K}}$ is an order ideal (or down-set) of $(2^{\mathcal{I}}, \subseteq)$ [21].

Proposition 3. (ANTI-MONOTONE CONSTRAINT) *Let g be an itemset. g fulfills the following two properties:*
1. If $g \in \mathcal{MG}\texttt{suc}_{\mathcal{K}}$, then $\forall\, g_1$ s.t. $g_1 \subset g$, $g_1 \in \mathcal{MG}\texttt{suc}_{\mathcal{K}}$.
2. If $g \notin \mathcal{MG}\texttt{suc}_{\mathcal{K}}$, then $\forall\, g_1$ s.t. $g \subset g_1$, $g_1 \notin \mathcal{MG}\texttt{suc}_{\mathcal{K}}$.

Proof.
1. $g \in \mathcal{MG}\texttt{suc}_{\mathcal{K}} \Longrightarrow \forall\, g_1$ s.t. $g_1 \subset g$, $g_1 \in \text{MG}\texttt{rep}_{f_1}$ with $f_1 = g_1''$ (According to Definition 5.) $\Longrightarrow \forall\, g_1$ s.t. $g_1 \subset g$, $g_1 \in \text{MG}\texttt{suc}_{f_1}$ (Since $\text{MG}\texttt{rep}_{f_1} \subseteq \text{MG}\texttt{suc}_{f_1}$.) $\Longrightarrow \forall\, g_1$ s.t. $g_1 \subset g$, $g_1 \in \mathcal{MG}\texttt{suc}_{\mathcal{K}}$ (Since $\text{MG}\texttt{suc}_{f_1} \subseteq \mathcal{MG}\texttt{suc}_{\mathcal{K}}$.).
2. $g \notin \mathcal{MG}\texttt{suc}_{\mathcal{K}} \Longrightarrow \forall\, g_1$ s.t. $g \subset g_1$, $g_1 \in \text{MG}\texttt{red}_{f_1}$ with $f_1 = g_1''$ (Indeed, g_1 has at least a non-representative subset, namely g, since the latter is not a succinct MG *and, hence, is not a representative one.) $\Longrightarrow \forall\, g_1$ s.t. $g \subset g_1$, $g_1 \notin \text{MG}\texttt{suc}_{f_1}$ (According to Definition 5, g_1 can not be redundant and succinct at the same time.) $\Longrightarrow \forall\, g_1$ s.t. $g \subset g_1$, $g_1 \notin \mathcal{MG}\texttt{suc}_{\mathcal{K}}$ (We have $g_1 \notin \text{MG}\texttt{suc}_{f_1}$. In addition, $g_1 \notin (\mathcal{MG}\texttt{suc}_{\mathcal{K}} \backslash \text{MG}\texttt{suc}_{f_1})$ since the closure of g_1 is unique and is equal to f_1.).* ♦

Since the frequency constraint is also anti-monotone, it is easy to show that the set $\mathcal{FMG}\texttt{suc}_{\mathcal{K}}$, of the *succinct frequent* MGs extracted from the context $\mathcal{K}$, is also an order ideal. This interesting property allowed us to propose an efficient algorithm to extract the SSMG according to the definition of Dong *et al.* (see [22] for more details).

3.2 Limitations of the work of Dong *et al.*

Starting form Definition 6, the main facts that can be pointed out from the work of Dong *et al.* can be unraveled by the following claims [20]:

Claim 1: The cardinality of a SSMG is insensitive to the considered total order relation $\preceq$, *i.e.*, whatever the total order relation, the number of *canonical* MGs is the same. Recall that the number of *representative* ones is exactly equal to that of CIs, as stated by Proposition 1.

Claim 2: A SSMG is an *exact representation* of the MG set, *i.e.*, if g is a *redundant* MG, then g can be inferred from the SSMG without loss of information. To do so, for each equivalence class, Dong *et al.* propose to infer its *redundant* MGs by replacing the subsets (one or more) of its *succinct* MGs by *non-representative* MGs having, respectively, the same closures as those of the replaced subsets [20]. For example, the *redundant* MG "*bdg*", extracted from the context sketched by Figure 1 (Left), can be inferred from the *succinct* MG "*adg*" by replacing its subset "*ad*" by "*bd*" (both MGs "*ad*" and "*bd*" have the same closure).

In what follows, we show that, according to the current definition of the SSMG, the cardinality of the latter closely depends on the selected total order relation (contrary to the statement of **Claim 1**). Furthermore, we give an example where the SSMG presents a loss of information (contrary to the statement of **Claim 2**).

As mentioned in the previous subsection, Dong *et al.* claimed that the shift of the total order relation $\preceq$ does not affect the size of the associated SSMG [20]. Such claim seems to be true when confronted to the extraction context depicted by Figure 1. Indeed, for different total order relations (*e.g.*, the lexicographic order, the ascending support order, the descending support order, etc.), we obtain the same number of *succinct* minimal generators (MGs). It is the same for the running example used in the proper paper of Dong *et al.* (see [20]). However, if we consider the extraction context sketched by Figure 2 (Left), we find that their claim is erroneous. Indeed, as shown by Figure 2 (Right), the total number of *succinct* MGs is equal to $\mathbf{23}$ if the lexicographic order is of use. Whereas, it is equal to $\mathbf{22}$ in the case of the ascending support order, and $\mathbf{25}$ in the case of the descending support order. Hence, the number of the *succinct* MGs closely depends on the chosen total order relation. The difference occurs within the equivalence class number $\mathbf{11}$ (shown with bold letters). The other equivalence classes do not contain any *redundant* MGs and, hence, are not of interest in our explanations.

Furthermore, if we adopt the ascending support order as a total order relation $\preceq$, then we find that, being given the *succinct* MGs, it is not possible to infer *all redundant* ones. Indeed, from the *succinct* MGs "*ea*" and "*acd*", only the two *redundant* MGs "*adf*" and "*cdf*" can be inferred by replacing the subset "*ac*" of "*acd*" by the MGs having its closure, *i.e.*, "*af*" and "*cf*". Hence, for example, the *redundant* MG "*edf*" will be missed if we need to infer *all* MGs.

Even if the first "bug" (*i.e.*, that related to the size of the different SSMGs associated to a given extraction context) can be regarded as not having a dramatic consequence, fixing the second one is of paramount importance, since the need for *exact* compact representation is always conditioned by the ability to discover *all redundant* information without looking back at the extraction context. Hence, aiming to ensure the completeness of the derivation of *redundant* MGs, we introduce, in the next section, new definitions allowing to go beyond the limitations of the work proposed by Dong *et al*.

3.3 Succinct System of Minimal Generators: new definitions

The set MG_f of the MGs associated to a given closed itemset (CI) f can be divided into different **equivalence subclasses** $^{(3)}$ thanks to an introduced substitution process. The latter uses a substitution operator denoted *Subst*. This substitution operator is a partial one allowing to substitute a subset of an itemset X, say Y, by another itemset Z belonging to the same equivalence class of Y (*i.e.*, $Y'' = Z''$). This operator is then defined as follows:

[3] The term *equivalence subclasses* is used here instead of *equivalence classes* to avoid the confusion with the *equivalence classes* induced by the closure operator $''$.

	a	*b*	*c*	*d*	*e*	*f*
1	×	×				
2	×		×			×
3	×			×		
4		×	×	×	×	
5	×	×	×	×	×	×
6			×	×		
7		×				×
8				×		×
9		×			×	×

	lexicographic order		***ascending support*** order		***descending support*** order	
	CI	**MGs**	**CI**	**MGs**	**CI**	**MGs**
1	∅	∅	∅	∅	∅	∅
2	*a*	***a***	*a*	***a***	*a*	***a***
3	*b*	***b***	*b*	***b***	*b*	***b***
4	*c*	***c***	*c*	***c***	*c*	***c***
5	*d*	***d***	*d*	***d***	*d*	***d***
6	*be*	***e***	*eb*	***e***	*be*	***e***
7	*f*	***f***	*f*	***f***	*f*	***f***
8	*ab*	***ab***	*ab*	***ab***	*ba*	***ba***
9	*acf*	***ac***, ***af***, ***cf***	*acf*	***ac***, ***af***, ***cf***	*fac*	***fa***, ***fc***, ***ac***
10	*ad*	***ad***	*ad*	***ad***	*da*	***da***
11	*abcdef*	***ae***, ***abc***, *abd*, *abf*, ***acd***, *adf*, *bcf*, *bdf*, *cdf*, *cef*, *def*	*eacbdf*	***ea***, *ecf*, *edf*, *acb*, ***acd***, *abd*, *abf*, *adf*, *cbf*, *cdf*, *bdf*	*bdface*	***ae***, ***bdf***, ***bda***, ***bfa***, *bfc*, *bac*, ***dfa***, *dfc*, *dfe*, *dac*, *fce*
12	*bcde*	***bc***, ***bd***, ***ce***, ***de***	*ecbd*	***ec***, ***ed***, ***cb***, ***bd***	*bdce*	***bd***, ***bc***, ***de***, ***ce***
13	*bf*	***bf***	*bf*	***bf***	*bf*	***bf***
14	*cd*	***cd***	*cd*	***cd***	*dc*	***dc***
15	*df*	***df***	*df*	***df***	*df*	***df***
16	*bef*	***ef***	*ebf*	***ef***	*bfe*	***fe***

Fig. 2. (**Left**) An extraction context $\mathcal{K}$. (**Right**) The CIs extracted from $\mathcal{K}$ and for each one, the corresponding MGs for different total order relations (the *succinct* MGs, according to the definition of Dong *et al.*, are indicated with bold letters).

Definition 7. (SUBSTITUTION OPERATOR) *Let* X, Y *and* Z *be three itemsets such that* $Y \subset X$ *and* $Y'' = Z''$*. The substitution operator* Subst*, w.r.t.* X, Y *and* Z*, is defined as follows :* Subst$(X, Y, Z) = (X \backslash Y) \cup Z$.

To prove that X and *Subst*(X, Y, Z) have the same closure, we need the following lemma.

Lemma 2. *[21] Let* X *and* Y *be two itemsets. X and Y verify the following property:* $(X \cup Y)'' = (X'' \cup Y'')''$.

In our context, with $X \cup Y$, we will indicate the ordered sequence of items, *w.r.t.* the total order relation $\preceq$, contained in X or in Y.

Proposition 4. X *and* Subst(X, Y, Z) *belong to the same equivalence class, induced by the closure operator* $''$.

Proof.
Let W *be the result of* Subst(X, Y, Z)*, i.e.,* $W = (X \backslash Y) \cup Z$*. We will show that* X *and* W *have the same closure.*
Using Lemma 2, we have: $X'' = ((X \backslash Y) \cup Y)'' = ((X \backslash Y)'' \cup Y'')''$*. Since* $Y'' = Z''$*, then* $X'' = ((X \backslash Y)'' \cup Y'')'' = ((X \backslash Y)'' \cup Z'')'' = ((X \backslash Y) \cup Z)'' = W''$*. Hence,* $X'' = W''$*. Thus, we can conclude that* X *and* W *necessarily belong to the same equivalence class, induced by the closure operator* $''$. ♦

For each equivalence class $\mathcal{C}$ (or equivalently, for each CI f), the substitution operator induces an equivalence relation on the set MG_f of the MGs of f portioning it into distinct equivalence subclasses. The definition of an equivalence subclass requires that we redefine the notion of *redundant* MG under the substitution process point of view. Indeed, according to the definition given by Dong *et al.* (see Definition 5), *redundant* MGs are blindly pruned according to purely syntactic properties, only consisting in checking the order of their subsets *w.r.t* $\preceq$, in their respective equivalence classes. Hence, we propose to incorporate a semantic part based on the actual concept of redundancy.

Definition 8. (MINIMAL GENERATORS' REDUNDANCY) *Let g and g_1 be two* MG*s belonging to the same equivalence class induced by the closure operator $''$.*

• *g is said to be a **direct redundant** (resp. derivable) with respect to (resp. from) g_1, denoted $g_1 \vdash g$, if Subst(g_1, g_2, g_3) = g with $g_2 \subset g_1$ and $g_3 \in \mathcal{MG}_{\mathcal{K}}$ s.t. $g_3'' = g_2''$. The operator $\vdash$ is reflexive, symmetric but not necessarily transitive.*

• *g is said to be a **transitive redundant** with respect to g_1, denoted $g_1 \vDash g$, if it exists a sequence of n MGs ($n \geq 2$), gen_1, gen_2, ..., gen_n, such that $gen_i \vdash gen_{i+1}$ ($i \in$ [1..$(n$-$1)$]) with gen_1 = g_1 and gen_n = g. The operator $\vDash$ is reflexive, symmetric and transitive.*

For n = 2, the operator $\vDash$ is reduced to the operator $\vdash$.

Definition 9. (EQUIVALENCE SUBCLASSES) *The operator $\vDash$ induces an equivalence relation on the set MG_f, of the* MG*s associated to a* CI *f, portioning it into distinct subsets called equivalence subclasses. If $g \in \text{MG}_f$, then the equivalence subclass of g, denoted by [g], is the subset of MG_f consisting of all elements that are transitive redundant w.r.t. g. In other words, we have: [g] = $\{g_1 \in \text{MG}_f \mid g \vDash g_1\}$.*
The smallest MG *in each equivalence subclass, w.r.t. the total order relation $\preceq$, will be considered as its **succinct*** MG*. While, the other* MG*s will be qualified as **redundant*** MG*s.*

The following pseudo-code offers a straightforward way to extract the different equivalence subclasses associated to a CI f. An equivalence subclass will be denoted *Equi_SubClass.*

Procedure 1: EQUIVALENCE SUBCLASSES MINER

Input: The set MG_f of the MGs associated to f.
Output: The equivalence subclasses associated to f.

Begin
  $\mathcal{S}$ = MG_f;
  i = 1;
  While $\mathcal{S} \neq \emptyset$ **do**
    g_s = $\min_{\preceq}(\mathcal{S})$; /*$g_s$ is the smallest MG in $\mathcal{S}$ *w.r.t.* $\preceq$.*/
    Equi_SubClass$_i$ = $\{g_s\} \cup \{g \in \mathcal{S} \mid g_s \vDash g\}$;
    $\mathcal{S}$ = $\mathcal{S} \backslash$*Equi_SubClass*$_i$;
    i = i + 1;
End

Example 3. Let us consider the extraction context depicted by Figure 2, the ascending support order as a total order relation $\preceq$ and the equivalence class having for CI "*eacbdf*". Using Procedure 1, the MGs associated to "*eacbdf*" are divided as follows:

1. First, $\mathcal{S}$ = MG$_{eacbdf}$ = {*ea*, *ecf*, *edf*, *acb*, *acd*, *abd*, *abf*, *adf*, *cbf*, *cdf*, *bdf*} and i = 1. "*ea*" is the *smallest* MG in $\mathcal{S}$. Hence, *Equi_SubClass*$_1$ = {*ea*} $\cup$ {$g \in \mathcal{S} \mid ea \vDash g$}. However, none MG can be deduced from "*ea*". Thus, *Equi_SubClass*$_1$ = {*ea*}.
2. Second, $\mathcal{S}$ = $\mathcal{S}\backslash$*Equi_SubClass*$_1$ = {*ea*, *ecf*, *edf*, *acb*, *acd*, *abd*, *abf*, *adf*, *cbf*, *cdf*, *bdf*}$\backslash${*ea*} = {*ecf*, *edf*, *acb*, *acd*, *abd*, *abf*, *adf*, *cbf*, *cdf*, *bdf*} and i = 2. "*ecf*" is the *smallest* one in $\mathcal{S}$. Hence, *Equi_SubClass*$_2$ = {*ecf*} $\cup$ {$g \in \mathcal{S} \mid ecf \vDash g$} = {*ecf*} $\cup$ {*edf*, *acb*, *abd*, *abf*, *cbf*, *bdf*}. Indeed, ***Subst***(ecf, *ec*, *ed*) = *edf* $\in$ MG$_{eacbdf}$ ($ecf \vdash$ *edf* and, hence, $ecf \vDash$ *edf*), ***Subst***(ecf, *ec*, *cb*) = *cbf* $\in$ MG$_{eacbdf}$ ($ecf \vdash$ *cbf* and, hence, $ecf \vDash$ *cbf*), ***Subst***(cbf, cf, *ac*) = *acb* $\in$ MG$_{eacbdf}$ ($ecf \vDash$ *acb* since $ecf \vdash$ *cbf* and then, $cbf \vdash$ *acb*), etc.
3. Finally, $\mathcal{S}$ = $\mathcal{S}\backslash$*Equi_SubClass*$_2$ = {*ecf*, *edf*, *acb*, *acd*, *abd*, *abf*, *adf*, *cbf*, *cdf*, *bdf*}$\backslash${*ecf*, *edf*, *acb*, *abd*, *abf*, *cbf*, *bdf*} = {*acd*, *adf*, *cdf*} and i = 3. "*acd*" is the *smallest* MG in $\mathcal{S}$. Hence, *Equi_SubClass*$_3$ = {*acd*} $\cup$ {$g \in \mathcal{S} \mid acd \vDash g$} = {*acd*} $\cup$ {*adf*, *cdf*} since ***Subst***(acd, ac, af) = *adf* ($acd \vdash adf$ and, hence, $acd \vDash adf$) and ***Subst***(acd, ac, cf) = *cdf* ($acd \vdash cdf$ and, hence, $acd \vDash cdf$).

In conclusion, MG$_{eacbdf}$ is divided into three equivalence subclasses as follows (*succinct* MGs are marked with bold letters): MG$_{eacbdf}$ = {***ea***} $\cup$ {***ecf***, *edf*, *acb*, *abd*, *abf*, *cbf*, *bdf*} $\cup$ {***acd***, *adf*, *cdf*}. Note that "***ecf***" was not considered as a succinct ***MG*** according to the initial definition that was introduced by Dong *et al.* since its subset "*cf*" is not the *representative MG* of its CI "*acf*". Hence, *all* MGs belonging to *Equi_SubClass*$_2$ can not be inferred according to their definition, contrary to ours.

Example 4. For the same context, if we consider the descending support order as a total order relation $\preceq$, then we will note that the SSMG, as formerly defined by Dong *et al.*, can even contain redundancy in comparison to our definition. Indeed, thanks to the substitution operator ***Subst***, MG$_{bdface}$ is divided as follows: MG$_{bdface}$ = {***ae***} $\cup$ {***bdf***, *bda*, *bfa*, *bfc*, *bac*, *dfe*, *fce*} $\cup$ {***dfa***, *dfc*, *dac*}. The storage of the MGs "*bda*" and "*bfa*" is then redundant and useless since they can simply be inferred starting from the *succinct* MG "*bdf*" (*bdf* $\vDash$ *bda* and *bdf* $\vDash$ *bfa*). Indeed, ***Subst***(*bdf*, *bd*, *bc*) = bfc, ***Subst***(bfc, fc, fa) = bfa, ***Subst***(bfa, fa, *ac*) = *bac* and finally ***Subst***(*bac*, *bc*, *bd*) = bda.

Property 1. The different equivalence subclasses associated to a given CI f verify the following properties:

- $\bigcup_{i=1}^{i \leq Card(\text{MGsuc}_f)}$ *Equi_SubClass*$_i$ = MG$_f$.
- $\forall\, i, j \in$ [1.. Card(MGsuc$_f$)] *s.t.* $i \neq j$, *Equi_SubClass*$_i$ $\cap$ *Equi_SubClass*$_j$ = $\emptyset$.

Using the new definitions of both *succinct* and *redundant* MGs (*c.f.*, Definition 8 and Definition 9), we can now define the succinct system of minimal generators (SSMG) in its new form as follows:

Definition 10. (SUCCINCT SYSTEM OF MINIMAL GENERATORS: NEW DEFINITION) *A succinct system of minimal generators (*SSMG*) is a system where only succinct* MG*s are retained among all* MG*s associated to each* CI.

Proposition 5. *Whatever the total order relation $\preceq$, the substitution operator* Subst *maintains unchanged the elements belonging to each equivalence subclass.*

Proof.
Let $\preceq_1$ and $\preceq_2$ be two different total order relations. Let f be a CI *and* MG_f *be the set of its associated* MG*s. Using $\preceq_1$,* MG_f *will be divided into equivalence subclasses. Let* Equi_Sub_Class$_{\preceq_1}$ *be one of them and g_{s_1} be its succinct* MG *(i.e., the smallest one in* Equi_Sub_Class$_{\preceq_1}$ *w.r.t. $\preceq_1$).* Equi_Sub_Class$_{\preceq_1}$ *can be represented by a tree, denoted $T_{\preceq_1}$. The root of $T_{\preceq_1}$ contains the succinct* MG *g_{s_1}. In this tree, a node N, which represents a* MG *g, points to a node N_1, which represents a* MG *g_1, if $g \vdash g_1$. Hence, from whatever node in $T_{\preceq_1}$, we can access the remaining nodes as follows: we move downward from the node N to the node N_1 using the relation $g \vdash g_1$ and conversely, from N_1 to N using the relation $g_1 \vdash g$. Indeed, if* Subst(g, g_2, g_3) = g_1 *with $g_2 \subset g$ and $g_3 \in \mathcal{MG}_{\mathcal{K}}$ s.t. $g_3'' = g_2''$, then also* Subst(g_1, g_3, g_2) = g *since the operator $\vdash$ is reflexive (cf. Definition 8).*

Now, consider the set Equi_Sub_Class$_{\preceq_1}$ *ordered w.r.t. to the second total order relation $\preceq_2$. The obtained new set will be denoted* Equi_Sub_Class$_{\preceq_2}$ *and its associated succinct* MG *will be denoted g_{s_2}. Hence, if we transform the tree $T_{\preceq_1}$ in a new one, denoted $T_{\preceq_2}$ and rooted in g_{s_2}, then we are able to reach all remaining* MG*s contained in* Equi_Sub_Class$_{\preceq_2}$ *thanks to the substitution application as explained above. Thus, the change of the total order relation does not affect the content of* Equi_Sub_Class$_{\preceq_1}$ *since it does not involve the deletion of any node in $T_{\preceq_1}$.*

Furthermore, this change does not augment Equi_Sub_Class$_{\preceq_2}$ *by any another redundant* MG*. Indeed, suppose that a* MG *denoted g_{new}, not already belonging to* Equi_Sub_Class$_{\preceq_1}$*, will be added to* Equi_Sub_Class$_{\preceq_2}$ *once we shift the total order relation from $\preceq_1$ to $\preceq_2$ (i.e., $g_{s_2} \vDash g_{new}$ but $g_{s_1} \nvDash g_{new}$). Since, $g_{s_1} \vDash g_{s_2}$ ($g_{s_2} \in$* Equi_Sub_Class$_{\preceq_1}$*) and $g_{s_2} \vDash g_{new}$, then $g_{s_1} \vDash g_{new}$ (according to Definition 8). Hence, g_{new} should belong to* Equi_Sub_Class$_{\preceq_1}$ *(according to Definition 9) what is in contradiction with the starting assumption ($g_1 \nvDash g_{new}$). Thus, $g_2 \nvDash g_{new}$.*

Therefore, we can conclude that the elements belonging to Equi_Sub_Class$_{\preceq_2}$ *are exactly the same than those contained in* Equi_Sub_Class$_{\preceq_1}$*, ordered w.r.t. $\preceq_2$ instead of $\preceq_1$.* ♦

Example 5. If we review Example 3 and Example 4, we note that *Equi_Sub_Class*$_1$, *Equi_Sub_Class*$_2$ and *Equi_Sub_Class*$_3$ are exactly the same for both examples. However, they are sorted according to the ascending support order and to the descending support order, respectively.

According to Proposition 5, the number of *succinct* MGs associated to each CI f (*i.e.*, Card($\text{MG}\mathtt{suc}_f$)) is then equal to the number of equivalence subclasses induced by the substitution operator, independently of the chosen total order relation. Hence, the cardinality of the set $\mathcal{MG}\mathtt{suc}_{\mathcal{K}}$, containing the *succinct* MGs that can be extracted from the context $\mathcal{K}$, remains unchanged even if we change the total order relation. In other words, the different SSMGs associated to an extraction context have the same size.

Proposition 6. *The* SSMG *as newly defined ensures the inference of each redundant* MG *g.*

Proof.
Since g is a redundant MG*, then g is not the smallest one in its equivalence subclass. Hence, according to the definition of an equivalence subclass (see Definition 9), it exists a succinct* MG *g_s belonging to the* SSMG *whose a substitution process certainly leads to g ($g_s \models g$) since the number of* MG*s belonging to each equivalence subclass is finite.*
♦

According to Proposition 6, the SSMG as newly defined becomes an *exact* representation of the MG set.

Proposition 5 and Proposition 6 make it possible to correct the claims of Dong *et al.* [20] thanks to the new semantic consideration of the concept of redundancy within the MG set.

4 Related work

In this part, we will mainly concentrate on the concept of **clone items** [23, 24] since it is closely related to our work. Clone items can be considered as a *restriction* of the SSMG to equivalence classes where two or more *items* have the same closure, *i.e.*, to MGs of size *one* (like the couple (*a*, *b*) and the couple (*d*, *e*) of the extraction context depicted by Figure 1 (Left)). The authors [23, 24] show that, for a couple like (*a*, *b*), items *a* and *b* present symmetries which can be seen as redundant information since for *all* association rules containing *a* in the antecedent there exists the same association rules where "*a*" is replaced by "*b*" [24]. Thus, they propose to ignore *all* rules containing "*b*" but not "*a*" without loss of information [24]. This reduction process was applied to the Guigues-Duquenne basis [25] for exact association rules. Association rules of this basis present implications between pseudo-closed itemsets [25] and closed itemsets. Note that clone items when applied to pseudo-closed itemsets are called *P-clone items* [23].

5 Experimental study

In order to evaluate the utility of our approach, we conducted series of experiments on four benchmark datasets, frequently used by the data mining community [(4)]. Characteristics of these datasets are summarized by Table 1. All experiments were run on a PC equipped with a 2.4GHz Pentium IV and 512MB of RAM. All programs were implemented in the C language. The operating system was S.U.S.E Linux 9.0 and we used gcc 3.3.1 for the compilation. Hereafter, we use a logarithmically scaled ordinate axis for all curves.

Figure 3 shows the effect of the succinct system of minimal generators (SSMG) by comparing the number of the *succinct frequent* minimal generators (MGs) to that of *all frequent* MGs. For the PUMSB and MUSHROOM datasets, a large part of the *frequent* MGs proves to be *redundant*. Indeed, for PUMSB (resp. MUSHROOM), in average **52.27%** (resp. **50.50%**) of the *frequent* MGs are *redundant*, and the maximum rate of redundancy reaches **64.06%** (resp. **53.28%**) for a *minsupp* value equal to **65%**

[4] These benchmark datasets are downloadable from: *http://fimi.cs.helsinki.fi/data*.

Dataset	**Number of items**	**Number of objects**	**Average object size**	***minsupp* interval (%)**
PUMSB	**7, 117**	**49, 046**	**74**	**90 - 60**
MUSHROOM	**119**	**8, 124**	**23**	**1 - 0.01**
CONNECT	**129**	**67, 557**	**43**	**90 - 50**
T40I10D100K	**1, 000**	**100, 000**	**40**	**10 - 1**

Table 1. Dataset characteristics.

(resp. $\mathbf{0.20\%}$). It is important to mention that for the PUMSB dataset, the redundancy is caused by the fact that there are some couples of items having the same closure (like "*a*" and "*b*" of the extraction context sketched by Figure 1 (Left)). Hence, using only an item, instead of both items forming each couple, was sufficient to eliminate all redundancy, which is not the case for the MUSHROOM dataset. Moreover, it is noteworthy that, in average, the number of *succinct* (resp. *all*) *frequent* MGs per equivalence class, is equal to $\mathbf{1.0004}$ (resp. $\mathbf{2.2382}$) for the PUMSB dataset, while it is equal to $\mathbf{1.0589}$ (resp. $\mathbf{2.1336}$) for the MUSHROOM dataset. Such statistics explain why the curve representing the number of *frequent* closed itemsets (CIs) is almost overlapped with that depicting the number of *succinct frequent* MGs.
For the CONNECT dataset and although it is known to be a "dense" one, each *frequent* CI extracted from this dataset has only a unique *frequent* MG and, hence, there are no *redundant* ones. It is the same for the "sparse" T40I10D100K dataset. Hence, it is worth noting that the reduction ratio from the number of *all frequent* MGs to that of *succinct* ones can be considered as a new measure for an improved dataset classification, as mentioned by Dong *et al.* [20].

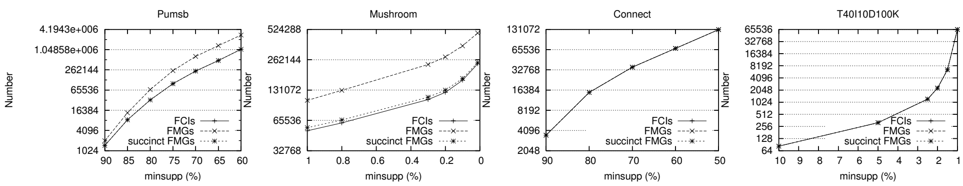

Fig. 3. The number of *frequent* CIs (denoted FCIs), of *frequent* MGs (denoted FMGs) and of *succinct frequent* MGs (denoted succinct FMGs).

Obtained results prove that the SSMG allows to almost reach the ideal case: $\mathbf{1}$ *succinct* MG per equivalence class.

6 Conclusion and future work

In this paper, we studied the principal properties of the succinct system of minimal generators (SSMG) as formerly defined by Dong *et al*. Once the limitations of the current definition mentioned, we introduced new ones aiming to make of the SSMG an *exact* representation of the minimal generator (MG) set, on the one hand, and, on the other hand, its size independent from the adopted total order relation. After that, we discuss the main related work to ours. Finally, an experimental study confirmed that the application of the SSMG makes it possible to get, in average, almost as many closed itemsets (CIs) as *succinct* MGs thank to the elimination of an important number of *redundant* ones.

As part of future work, we plan to use of the SSMG in an in-depth structural analysis of dataset characteristics. In this context, we will propose a sparseness measure based on the SSMG. The extension of the SSMG to the framework of generic association rules is also an interesting issue. As a first attempt, the work we proposed in [26] gave very encouraging results. Furthermore, we think that the application of the SSMG to some real-life domains like biological applications will be of an added value for the end-users.

Acknowledgements

We are grateful to the anonymous reviewers for their useful remarks and suggestions. The first author also thanks Petko Valtchev and Christian Pagé for their helpful comments on a previous version of this paper. This work is partially supported by the French-Tunisian project CMCU 05G1412.

References

1. Bastide, Y., Taouil, R., Pasquier, N., Stumme, G., Lakhal, L.: Mining frequent patterns with counting inference. In: Proceeding of the 6th ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining, volume 2(2), Boston, Massachusetts, USA. (2000) 66–75
2. Bastide, Y., Pasquier, N., Taouil, R., Stumme, G., Lakhal, L.: Mining minimal non-redundant association rules using frequent closed itemsets. In: Proceedings of the 1st International Conference on Computational Logic (DOOD 2000), Springer-Verlag, LNAI, volume 1861, London, UK. (2000) 972–986
3. Boulicaut, J.F., Bykowski, A., Rigotti, C.: Free-sets: A condensed representation of boolean data for the approximation of frequency queries. In Jounal of Data Mining and Knowledge Discovery (DMKD), volume 7(1). (2003) 5–22
4. Stumme, G., Taouil, R., Bastide, Y., Pasquier, N., Lakhal, L.: Computing iceberg concept lattices with TITANIC. Journal on Knowledge and Data Engineering (KDE), volume 2 (42). (2002) 189–222
5. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. Journal of Information Systems, volume 24 (1) (1999) 25–46
6. Pasquier, N., Bastide, Y., Touil, R., Lakhal, L.: Discovering frequent closed itemsets for association rules. In Beeri, C., Buneman, P., eds.: Proceedings of 7th International Conference on Database Theory (ICDT 1999), LNCS, volume 1540, Springer-Verlag, Jerusalem, Israel. (1999) 398–416
7. Li, H., Li, J., Wong, L., Feng, M., Tan, Y.: Relative risk and odds ratio: a data mining perspective. In: Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART symposium on Principles Of Database Systems (PODS 2005), Baltimore, Maryland, USA. (2005) 368–377
8. Lucchese, C., Orlando, S., Perego, R.: Fast and memory efficient mining of frequent closed itemsets. In IEEE Journal Transactions on Knowledge and Data Engineering (TKDE) **18 (1)** (2006) 21–36
9. Pei, J., Han, J., Mao, R.: CLOSET: An efficient algorithm for mining frequent closed itemsets. In: Proceedings of the ACM-SIGMOD International Workshop on Data Mining and Knowledge Discovery (DMKD 2000), Dallas, Texas, USA. (2000) 21–30
10. Uno, T., Asai, T., Uchida, Y., Arimura, H.: An efficient algorithm for enumerating closed patterns in transaction databases. Proceedings of the 7th International Conference on Discovery Science, Padova, Italy (2004) 16–31

11. Zaki, M.J., Hsiao, C.J.: CHARM: An efficient algorithm for closed itemset mining. In: Proceedings of the 2nd SIAM International Conference on Data Mining, Arlington, Virginia, USA. (2002) 34–43
12. Ben Yahia, S., Hamrouni, T., Mephu Nguifo, E.: Frequent closed itemset based algorithms: A thorough structural and analytical survey. In: Proceedings of ACM-SIGKDD Explorations, volume 8(1). (2006) 93–104
13. Lucchese, C., Orlando, S., Palmerini, P., Perego, R., Silvestri, F.: kDCI: A multi-strategy algorithm for mining frequent sets. In Goethals, B., Zaki, M.J., eds.: Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI 2003), volume 90 of CEUR Workshop Proceedings, Melbourne, Florida, USA. (2003)
14. Hamrouni, T., Ben Yahia, S., Slimani, Y.: PRINCE: An algorithm for generating rule bases without closure computations. In Tjoa, A.M., Trujillo, J., eds.: Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DAWAK 2005), Springer-Verlag, LNCS, volume 3589, Copenhagen, Denmark. (2005) 346–355
15. Kryszkiewicz, M.: Concise representation of frequent patterns and association rules. Habilitation dissertation, Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland. (2002)
16. Zaki, M.J.: Mining non-redundant association rules. Journal of Data Mining and Knowledge Discovery (DMKD), volume 9. (2004) 223–248
17. Gasmi, G., BenYahia, S., Mephu Nguifo, E., Slimani, Y.: $\mathcal{IGB}$: A new informative generic base of association rules. In: Proceedings of the International 9th Pacific-Asia Conference on Knowledge Data Discovery (PAKDD 2005), Springer-Verlag, LNAI, volume 3518, Hanoi, Vietnam. (2005) 81–90
18. De Raedt, L., Ramon, J.: Condensed representations for inductive logic programming. In Dubois, D., Welty, C., eds.: Proceedings of the Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR 2004), Whistler, Canada. (2004) 438–446
19. Zhao, L., Zaki, M.J., Ramakrishnan, N.: BLOSOM: A framework for mining arbitrary boolean expressions over attribute sets. Technical report (http://www.cs.rpi.edu/research/), Department of Computer Science, Rensselaer Polytechnic Institute, USA. (2006)
20. Dong, G., Jiang, C., Pei, J., Li, J., Wong, L.: Mining succinct systems of minimal generators of formal concepts. In: Proceedings of the 10th International Conference on Database Systems for Advanced Applications (DASFAA 2005), Springer-Verlag, LNCS, volume 3453, Beijing, China. (2005) 175–187
21. Ganter, B., Wille, R.: Formal Concept Analysis. Springer-Verlag (1999)
22. Hamrouni, T., Ben Yahia, S., Mephu Nguifo, E.: Redundancy-free generic bases of association rules. In: Proceedings of the 8th French Conference on Machine Learning (CAp 2006), Presses Universitaires de Grenoble, Trégastel, France. (2006) 363–378
23. Gely, A., Medina, R., Nourine, L., Renaud, Y.: Uncovering and reducing hidden combinatorics in Guigues-Duquenne bases. In Ganter, B., Godin, R., eds.: Proceedings of the 3rd International Conference on Formal Concept Analysis (ICFCA 2005), Springer-Verlag, LNCS, volume 3403, Lens, France. (2005) 235–248
24. Medina, R., Nourine, L., Raynaud, O.: Interactive association rules discovery. In Missaoui, R., Schmidt, J., eds.: Proceedings of the 4th International Conference on Formal Concept Analysis (ICFCA 2006), Springer-Verlag, LNCS, volume 3874, Dresden, Germany. (2006) 177–190
25. Guigues, J.L., Duquenne, V.: Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines. **24** (1986) 5–18
26. Hamrouni, T., Ben Yahia, S., Mephu Nguifo, E.: Generic association rule bases: Are they so succinct? In: Proceedings of the 4th International Conference on Concept Lattices and their Applications (CLA 2006), Hammamet, Tunisia. (2006)