STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LV**, Number 1, 2010

DISCOVERING PATTERNS IN MUSIC. AN FCA GROUNDED APPROACH

ZSUZSANNA MARIAN AND CHRISTIAN SĂCĂREA

Abstract. We apply methods of Formal Concept Analysis in order to discover patterns in modern music. For this, we investigate the *Music Genome*, a project started on the 6th of January 2000 by a group of musicians and music-loving technologists who came together with the idea of creating the most comprehensive analysis of music ever. We also present an environment for processing and representation of music patterns.

1. Introduction

Motto: How sweet the moonlight sleeps upon this bank! Here will we sit, and let the sound of music Creep in our ears; soft stillness and the night Become the touches of sweet harmony.

– Shakespeare (Merchant of Venice, Lorenzo, Act 5, scene 1)

Discovering patterns in music is a usual research topic in music theory (see for instance [4], [5], and [6]). Also, the connection between music and mathematics and physics has been intensively studied along the centuries (see also [2], [3]).

Nevertheless, there are surprisingly few attempts to apply modern knowledge processing and representation methods to the study of musical patterns. The only attempt we know was an experiment conducted by a group of mathematicians and musicians at the Darmstadt University of Technology, Germany. An exhaustive attribute exploration was performed on a set of attributes for classical music in order to obtain a knowledge universe, i.e., a complete set of examples and counterexamples related to the considered set of attributes.

Received by the editors: January 18, 2010.

2010 *Mathematics Subject Classification.* 68P15 Database theory, 03G10 Lattices and related structures.

1998 *CR Categories and Descriptors.* code H.2.1 [**Database Management**]: Logical Design – *Normal Forms*; F.4.m [**Mathematical Logic and Formal Languages**]: Miscellaneous.

Key words and phrases. Conceptual Knowledge Processing, Formal Concept Analysis, Music Genome.

In this paper, we have performed a first study on music patterns for modern music. For this, we have used the so-called music genes, essential attributes for melodies, which are then combined to music. Like the discovery of the human genome, the discovery of the music genome was the result of an intensive work of a group of enthusiastic musicians and music loving technologists. The musical pool under investigation was modern music. While classical music has been exhaustively studied, modern music is actually a huge field for research, insufficiently exploited by modern science.

The French-born composer Edgar Varese (1883–1965) once defined music to be organized sound. But how is music organized and what type of organization is intended?

Classical music theory gives a partial answer to these questions, in terms of tones and harmonies. The Music Genome Project goes one step ahead and investigates the building harmonies of melodies, the so-called genes.

But how are these genes related? How are they related to the music in which they appear, what is the knowledge they encode? How does the logic of these genes look like and how can we discover as much knowledge as possible by studying the Music Genome?

Conceptual Knowledge Processing and its mathematical theory on which this is grounded, the Formal Concept Analysis, appears to open the possibility not only for a satisfactory answer to the questions above, but also for a comprehensive analysis of data and related concepts.

There is a long philosophical tradition in investigating concepts, ordering them in a certain hierarchy of subconcept-superconcept. This tradition has been reflected several times in the development of mathematics, and it can be refound in almost every part of modern mathematics: remember the nineteenth-century attempts to formalize logic (see for example the *Algebra of Logic* of Ernst Schröder). Traditionally, a concept is determined by its extent and its intent (or comprehension). The extent of a concept consists of all objects, individuals or entities which belong to the concept, while the intent consists of all properties which are considered valid for that concept. The hierarchy of concepts is given by the relation of subconcept wrt. certain superconcept, i.e., the extent of a subconcept is part of the extent of the superconcept, while the inverse relation holds for the corresponding intents.

Consider a usual concept with which one usually deal, for example landscape. There is no possibility to list every object of this concept or even to consider in a generally satisfactory way all properties of that what we understand as a landscape. It follows that there is an intrinsic necessity to restrict ourselves to a given set of objects and a clearly defined set of attributes, process which is characteristic for the human thinking, cutting off sections of the reality which will be here later on called formal contexts.

Formal Concept Analysis was born from this vigorous philosophical tradition. At the end of the 1970s, Formal Concept Analysis has crystalized first in an attempt of restructuring lattice theory (see [14]), where R. Wille discusses the status of modern lattice theory considering that "...abstract developments should be brought back to the commonplace in perception, thinking, and action. Thus *restructuring lattice theory* is understood as an attempt to unfold lattice-theoretical concepts, results, and methods in a continuous relationship with their surroundings."

As a mathematical theory, Formal Concept Analysis is based on the formalization of the notion of concept and of the medium from where this concept arises, the formal context. Formally speaking a formal context (mathematically defined in the next section) is a triple consisting of two sets and a binary relation between them. Despite of its simplicity, a formal context encodes in its incidence relation some structural information which can be found in the so-called formal concepts, which can be seen as a kind of closed pieces of the information encoded in the considered formal context.

From now on, Formal Concept Analysis became more as an attempt of reconsidering parts of lattice theory, giving the possibility of investigating a wide range of empirical phenomena using mathematical methods in the study of the data collected about these phenomena. FCA is not only a part of Mathematics, it has deep going results and connections with Computer Science, Medical Science, Social Sciences, Psychology and others.

2. Formal Concept Analysis

2.1. **Context and Concept.** As we have seen before, Formal Concept Analysis is based on a set theoretical model proposing a new paradigm of thinking. A **formal context** $\mathbb{K} := (G, M, I)$ consists of two sets G and M and a binary relation I between G and M. The elements of G are called **objects** (in German *Gegenstände*) and the elements of M are called **attributes** (in German *Merkmale*). The relation I is called the incidence relation of the formal context, and we sometimes write gIm instead of $(g, m) \in I$. If gIm holds, we say that *the object g has the attribute m.*

A small context is usually represented by a cross table, i.e., a rectangular table of crosses and blanks, where the rows are labeled by the objects and the columns are labeled by the attributes. A cross in entry (g, m) indicates gIm. Figure 1 illustrates a formal context. The attributes are the so-called *acoustic instruments music genes*, i.e., Subtle Use of Acoustic Piano, Acoustic Rhythm Guitars, Good Dose of Acoustic Guitar Pickin, Acoustic Guitar Riffs, Acoustic Rhythm Piano, Use of Acoustic Piano, Guitar, Piano. The objects are songs, having in their structure at least one of the considered music genes. Since the list of songs is very large, we have displayed as object names only the number of songs into consideration. The incidence relation is displayed by a series of crosses in the table, indicating when a specific song has a given music gene.

A	B	C	D	E	F	G	H	I
	Subtle Use of Acoustic Piano	Acoustic Rhythm Guitars	Good Dose of Acoustic Guitar Pickin	Acoustic Guitar Riffs	Acoustic Rhythm Piano	Use of Acoustic Piano	Guitar	Piano
18	×					×		×
130		×					×	
20			×				×	
84					×			×
1						×		×
3				×			×	

FIGURE 1. The Acoustic Instruments context.

For a set $A \subseteq G$ of objects we define

$$A' := \{m \in M \mid gIm \text{ for all } g \in A\}$$

the set of all attributes common to the objects in A. Dually, for a set $B \subseteq M$ of attributes we define

$$B' := \{g \in G \mid gIm \text{ for all } m \in B\}$$

the set of all objects which have all attributes in B.

A **formal concept** of the context $\mathbb{K} := (G, M, I)$ is a pair (A, B) where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$. We call A the **extent** and B the **intent** of the concept (A, B). The set of all concepts of the context (G, M, I) is denoted by $\mathfrak{B}(G, M, I)$.

2.2. **Many-valued contexts.** As we have seen from the definition of a formal context, an object could have or not some attribute, i.e., the attributes were one-valued. The more general situation is that where attributes have values. We call them *many-valued attributes*, in contrast to the *one-valued attributes* considered so far.

Definition 1. A **many-valued context** (G, M, W, I) consists of sets G, M and W and a ternary relation I between G, M and W (i.e., $I \subseteq G \times M \times W$) for which the following holds:

$$(g, m, w) \in I \text{ and } (g, m, v) \in I \text{ always imply } m = v.$$

The elements of G are called **objects**, those of M (**many-valued**) **attributes** and those of W **attribute values**.

Like the one-valued contexts treated so far, many-valued contexts can be represented by tables, the rows of which are labelled by the objects and the columns labelled by the attributes. The entry in row g and column m represents then the attribute value $m(g)$. If the attribute m does not have a value for the object g, there will be no entry.

In order to assign concepts to a many-valued context we first have to transform it into a formal context, according to some rules, rules which are called **scaling the many-valued context**. The concepts of this *derived* one-valued context are then interpreted as the concepts of the many-valued context. Since data captured in a many-valued context can be very complex, it is obviously that the interpretation process through which a conceptual structure is assigned, called *conceptual scaling* (see [GW 99]) is not uniquely determined. Changing the scale we can get new insights about the information of a given many-valued context. In the process of scaling, first of all each attribute of a many-valued context is interpreted by means of a context. This context is called *conceptual scale*.

Definition 2. A **scale** for the attribute m of a many-valued context is a (formal) context $\mathbb{S}_m := (G_m, M_m, I_m)$ with $m(G) \subseteq G_m$. The objects of a scale are called **scale values**, the attributes are called **scale attributes**.

As we can see, there is no formal difference between a formal context and a scale. But according to the tradition of the Darmstadt school of Formal Concept Analysis, the term *scale* will be used only for contexts which have a clear conceptual structure and which bear meaning.

There are several ways in which a scale for a given attribute $m \in M$ can be chosen. The same holds for combining the scales in order to obtain a formal context from the given many-valued one. The simplest case is called **plain scaling**, and it consists of putting together the individual scales without connecting them.

Definition 3. If (G, M, W, I) is a many-valued context and $\mathbb{S}_m$, with $m \in M$, are scales, then the **derived context with respect to plain scaling** is the context (G, N, J) with

$$N := \bigcup_{m \in M} \{m\} \times M_m,$$

and

$$gJ(m, n) :\Leftrightarrow m(g) = w \text{ and } wI_m n.$$

3. The Music Genome

The Music Genome Project was started on the 6th of January 2000 by a group of musicians and music-loving technologists who came together with the idea of creating the most comprehensive analysis of music ever, according to Tim Westergren, one of the founders of this project [7]. They have decided to analyze the structure of a song, defining different attributes - "genes" - for them and identify similar songs that the listener could be interested in. Nolan Grasser, an actual musicologist whose doctoral thesis dealt with the close analysis of Renaissance composition, helped Westergren to create the lexicon that could transform his genome idea into something a computer could evaluate [8]. During the analysis, every song is first broken down into the large-scale aspects of the music: melody, harmony, rhythm, form, sound and in many cases the lyrics. Each of these categories might have 10, 30, 50 elements. This is how a melody ends up being described by about 400 different genes.

This analysis of the songs is made by a team of analysts, musicians who have studied music for at least 4 years and have passed a music theory exam and completed 40 hours of training. This training was developed with the help of Nolan and is designed to make sure that the analysts are consistent with subjective matters, for example: how "emotionally intense" on a scale from 1 to 5 is the solo part in a given song [8]. Even with the training, about 10% of the songs are analyzed a second time by a senior analyst and any difference in the opinion over a genes is flagged and reviewed [9].

Due to the time requirements of the analysis (20 to 30 minutes for a 4 minute long song), 5 years and 30 music theory experts were needed to built a database that can be useful. In May 2006, this database contained over 400000 analyzed songs from more than 20000 contemporary artists [10] which was increased to 700000 songs from more than 80000 artists till October 2009 with the estimation that about 10000 new song are added monthly to the database [8].

For this database, an interface was created and made available online under the name Pandora Internet Radio. The Pandora player chooses a succession of songs from its database after the user creates a channel. This is done by entering the name of a song or an artist and hitting the create button. Then Pandora sorts a selection based on the characteristics of the song or artist [11]. This selection can be refined by pushing the buttons "Thumb up" and "Thumb down" on this interface, thus showing if the melody that is being played is liked or disliked. Moreover, it has a link saying: "Why did you play this song?" which takes the listener to a page where some of the genetic elements of the song are presented.

This approach of using the genes of a melody to find similar ones is a novel one, compared to the traditional recommendation systems of the type: "customers who liked this item also liked that one", and ignoring the crowd, saying that the taste of your cool friends, your peers, the traditional music critics, big-label talent scouts and the latest influential music blog are all equally irrelevant, because that is cultural information not musical one. To minimize the cultural influence during song listening, Westergren initially wanted to hide the artist of the songs until the listener asked to see it, but in the end he abandoned this idea [8].

Unfortunately, due to licensing problems, the Pandora Internet Radio is available only in the United States. Still, the search part (for artists and/or songs) of the site is available everywhere, and for many songs some of the "genes" are presented too, probably the same ones that a United States user sees when presses the "Why did you play this song?" link. The number of genes varies from song to song, usually there are between 4-20.

For this project, only the genes available at the Pandora site were used (there are approximative 150 such genes describing alternative rock melodies), since more specific information is considered proprietary information and cannot be publically released.

Since Pandora has a huge number of songs in its database, we only investigate songs of type "Alternative Rock" and from albums that appeared between 2007 and 2009. The list of all alternative rock artists is available at Wikipedia [12], this is where the name of artists were taken, searched with Pandora, and for the melodies from albums between 2007 and 2009 the available genes were recorded into a formal context together with the title, artist, album and year the album appeared. Despite the limitations in style and time, there still are a huge number of songs, so, due to lack of time, the context is not complete yet. It contains 451 songs, from 69 albums from 47 artists, which means the beginning of the list of artists whose name starts with "C" in the list at Wikipedia. As specified before, we had to collect the data for this analysis over the Internet, since more information is considered proprietary information and could not be released for our research.

This data resulted in a context with 451 objects (the songs) and 173 attributes (the genes). Some of these attributes are many-valued (for example most attributes related to the vocal characteristic of a song are many valued: they can have the value male of female) while some are one-valued. To scale this context and to create different concept lattices, the Elba editor of the ToscanaJ [13] application was used. To do so, the context was put in an Access database table, to which Elba can connect. Because of the dimension of the data table, the visualization of the entire concept lattice is not very practical, since it has a very complex structure.

We have decided to focus on subsets of attributes, selecting subcontexts and investigating the corresponding conceptual hierarchies. Similar attributes have been grouped together, creating scales and lattices for these. This gave 23 different scales such as: Acoustic Instruments, Electric Instruments, Lyrics, Vocals, Roots, Influences and others. Unfortunately there are 31 attributes which does not appear in either of these scales, but there are attributes which appear in more than one (for example the attribute Acoustic Piano belongs to the concept lattices Acoustic Instruments and Piano).

4. Investigating the Music Genome

As stated above, we have started the analysis of the Music Genome using specific methods of Conceptual Knowledge Processing. For this, data was gathered in tables, named contexts, then several subcontexts were considered, covering topics of interest for this study. Conceptual hierarchies have been build, and a detailed discussion of the results has been performed in order to highlight the conceptual connections between genes and melodies.

The attribute logic has been investigated by means of computing the stem base for attribute implications for each of the subcontexts. Also, association rules have been mined.

In the following, we illustrate these methods on 3 different subcontexts.

4.1. **Acoustic Instruments Context.** There are 5 attributes that belong to the acoustic context: Acoustic Rhythm Piano, Use of Acoustic Piano, Acoustic Guitar Riffs, Acoustic Rhythm Guitars, Good Dose of Acoustic Guitar Pickin'. Only one of these attributes, the Use of Acoustic Piano is multivalued, it appears in the context either simply "Use of Acoustic Piano" (1 object) or "Subtle Use of Acoustic Piano" (18 objects). Since it is a multivalued attribute we need a scale to transform it. An Ordinal Scale has been chosen, since Subtle Use of Acoustic Piano is still Use of Acoustic Piano. Moreover, two more attributes have added, Piano and Guitar to show the two main categories the acoustic instruments belong to and so to highlight the background knowledge on acoustic instruments (see [1]).

The reduced context that contains only these attributes has 187 objects, and its concept lattice can be seen on Figure 2. The nodes on the figure represent concepts, the higher a node is, the more general the concept is. Normally, all nodes connected directly to the uppermost node should be placed on the same level, and so on, but this is impossible because of the high number of nodes and the length of the labels. Yet, it is clearly visible that the concept with label "Use of Acoustic Piano" is more general than the one with label "Subtle Use of Acoustic Piano", because the latter is below the former in the lattice. The numbers associated to the nodes represent the number of objects in the context that have the given attribute, and not the total number of objects that have that attribute or one of its more specific ones. This is why the above mentioned "Subtle Use of Acoustic Piano" has the number 18 attached and the "Use of Acoustic Piano" has only the number 1. The uppermost node represents the attributes that all objects have (in our case there is no such object), while the lower one represents the objects that have all the attributes (in our case this is also empty).

Figure 3 displays the stem base for the 8 attributes of the Acoustic Instruments subcontext.

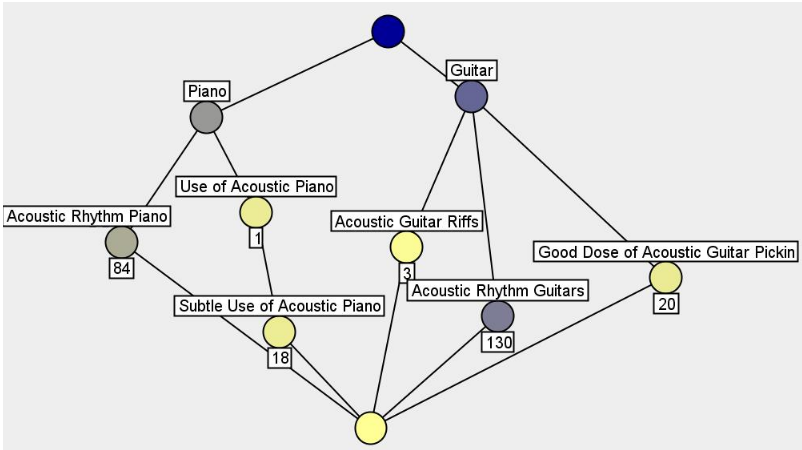

FIGURE 2. Concept lattice for attributes belonging to the Acoustic Instruments subcontext.

```
1 < 1 > Acoustic Rhythm Guitars ==> Guitar;
2 < 1 > Good Dose of Acoustic Guitar Pickin ==> Guitar;
3 < 1 > Acoustic Guitar Riffs ==> Guitar;
4 < 0 > Guitar Piano ==> Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Acoustic Rhythm Piano Use of Acoustic Piano;
5 < 1 > Subtle Use of Acoustic Piano ==> Use of Acoustic Piano Piano;
6 < 1 > Acoustic Rhythm Piano ==> Piano;
7 < 2 > Use of Acoustic Piano ==> Piano;
8 < 0 > Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Guitar ==> Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Acoustic Rhythm Piano Use of Acoustic Piano Piano;
9 < 0 > Acoustic Rhythm Guitars Acoustic Guitar Riffs Guitar ==> Subtle Use of Acoustic Piano Good Dose of Acoustic Guitar Pickin Acoustic Rhythm Piano Use of Acoustic Piano Piano;
10 < 0 > Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Guitar ==> Subtle Use of Acoustic Piano Acoustic Guitar Riffs Acoustic Rhythm Piano Use of Acoustic Piano Piano;
11 < 0 > Acoustic Rhythm Piano Use of Acoustic Piano Piano ==> Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Guitar;
```

FIGURE 3. List of implications for the attributes in the Acoustic instruments subcontext.

Remark 1. This base contains 11 implications 6 written in blue and 5 in red. The blue ones represent implications that have objects in the context that support that rule premise. These 6 implications are obvious, they follow directly from the way the scales for this context were defined. For example the fifth implication (Subtle Use of Acoustic Piano $\rightarrow$ Use of Acoustic Piano, Piano) follows from the ordinal scale we used for scaling the many-valued attribute. The red color marks implications which have no objects in the context that support the implication. This usually means that the set of objects from the premise does not occur together in the context. Such implications usually contains all the attributes. For example the first red implication is: Guitar, Piano $\rightarrow$ Subtle Use of Acoustic Piano, Acoustic Rhythm Guitars, Good Dose of Acoustic Guitar Pickin, Acoustic Guitar Riffs, Acoustic Rhythm Piano, Use of Acoustic Piano; so, the first two attributes imply all the rest. This obviously is an implication without objects because there is no instrument which would be Guitar and Piano in the same time.

Figure 4 displays a list of association rules with minimal support 0 and confidence 30%.

```
1 < 1 > Subtle Use of Acoustic Piano =[100%]=> < 1 > Use of Acoustic Piano Piano;
2 < 0 > Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Guitar =[100%]=> < 0 > Subtle Use of Acoustic Piano Acoustic Guitar Riffs Acoustic Rhythm Piano Use of Acoustic Piano Piano;
3 < 0 > Acoustic Rhythm Guitars Acoustic Guitar Riffs Guitar =[100%]=> < 0 > Subtle Use of Acoustic Piano Good Dose of Acoustic Guitar Pickin Acoustic Rhythm Piano Use of Acoustic Piano Piano;
4 < 1 > Acoustic Rhythm Guitars =[100%]=> < 1 > Guitar;
5 < 0 > Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Guitar =[100%]=> < 0 > Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Acoustic Rhythm Piano Use of Acoustic Piano Piano;
6 < 1 > Good Dose of Acoustic Guitar Pickin =[100%]=> < 1 > Guitar;
7 < 1 > Acoustic Guitar Riffs =[100%]=> < 1 > Guitar;
8 < 0 > Acoustic Rhythm Piano Use of Acoustic Piano Piano =[100%]=> < 0 > Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Guitar;
9 < 1 > Acoustic Rhythm Piano =[100%]=> < 1 > Piano;
10 < 2 > Use of Acoustic Piano =[100%]=> < 2 > Piano;
11 < 0 > Guitar Piano =[100%]=> < 0 > Subtle Use of Acoustic Piano Acoustic Rhythm Guitars Good Dose of Acoustic Guitar Pickin Acoustic Guitar Riffs Acoustic Rhythm Piano Use of Acoustic Piano;
12 < 3 > Piano =[67%]=> < 2 > Use of Acoustic Piano;
13 < 6 > { } =[50%]=> < 3 > Piano;
14 < 6 > { } =[50%]=> < 3 > Guitar;
15 < 2 > Use of Acoustic Piano Piano =[50%]=> < 1 > Subtle Use of Acoustic Piano;
16 < 3 > Piano =[33%]=> < 1 > Acoustic Rhythm Piano;
17 < 3 > Guitar =[33%]=> < 1 > Acoustic Guitar Riffs;
18 < 3 > Guitar =[33%]=> < 1 > Good Dose of Acoustic Guitar Pickin;
19 < 3 > Guitar =[33%]=> < 1 > Acoustic Rhythm Guitars;
```

FIGURE 4. List of association rules for the attributes in the Acoustic instruments subcontext.

Remark 2. There are 19 association rules out of which the first 11 correspond to the 11 implications from the other Figure, which is normal, since every implication is an association rule with a 100% support. The red color shows again association rules with no objects, blue color the association rules which are also implications (they have 100% support) while the green rules represents the so called *not strict rules* which occur only for some percent of the objects (their support is less than 100%). The support of every rule is immediately after the premise of the rule, between square brackets. Rules number 13 and 14 have as premise {}, which means that it holds for every object from the context. These show that there is a 50% chance that an object has the Piano or Guitar attribute. For example, the last three implications show that if an song has the attribute Guitar there is a 33% chance that is has either Acoustic Guitar Riffs or Good Dose of Acoustic Guitar Pickin or Acoustic Rhythm Guitars. This is true, because the attribute Guitar appears only if one of these three attribute is present. On the other hand, probably the combinations of the three attributes have a too small support.

4.2. **Electric Instruments Context.** The Electric Instruments context, although similar to the acoustic one, is a little more complex, since from the original context 7 attributes have been considered in this group: Electric Guitars, Electric Guitar Effects, Electric Guitar Riffs, Electric Guitar Solo, Electric Guitar Wall-o-sound, Electric Rhythm Guitars and Electric Pianos. The multi-valued attributes are the following: Electric Guitar Riffs (with values Electric Guitar Riffs and Dirty Electric Guitar Riffs), Electric Rhythm Guitars (with values Electric Rhythm Guitars and Heavy Electric Rhythm Guitars) and Electric Guitar Solo (with values of Electric Guitar Solo and Dirty Electric Guitar Solo). In all three cases an Ordinal scale was used. Besides this, all objects with attributes Electric Rhythm Guitars and Heavy Electric Rhythm Guitars were also assigned the attribute Electric Guitars because it seemed logic to do so. Similarly to the case with the acoustic instruments two more attributes (Guitar and Piano) were introduced in order to express a certain amount of background knowledge, but in this case there is only one attribute that falls in the Piano category (Electric Pianos) this is why that node has two labels. This sub-context contains 266 objects and the corresponding concept lattice can be seen on Figure 5.

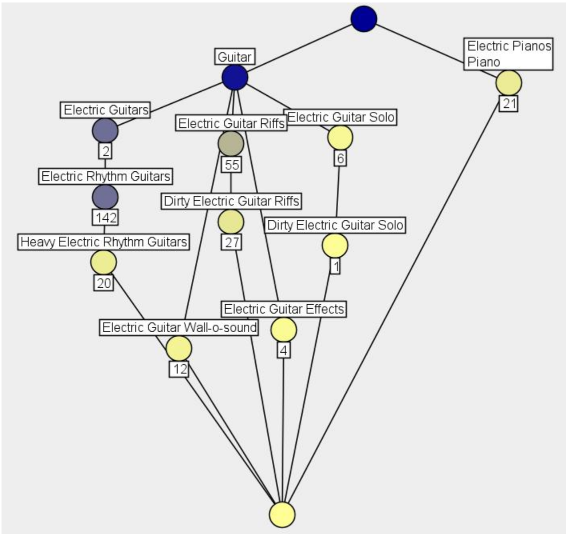

FIGURE 5. Concept lattice for attributes belonging to the Electric Instruments subcontext.

Figure 6 displays the stem base for the 11 attributes of the Electric Instruments subcontext.

1 < 1 > Electric Guitar Wall-o-sound ==> Guitar;
2 < 1 > Electric Guitar Effects ==> Guitar;
3 < 1 > Heavy Electric Rhythm Guitars ==> Electric Rhythm Guitars Electric Guitars Guitar;
4 < 2 > Dirty Electric Guitar Solo ==> Guitar;
5 < 2 > Electric Rhythm Guitars ==> Electric Guitars Guitar;
6 < 2 > Electric Guitar Riffs ==> Guitar;
7 < 2 > Electric Guitar Solo ==> Guitar;
8 < 3 > Electric Guitars ==> Guitar;

FIGURE 6. List of implications for the attributes in the Electric instruments subcontext.

Remark 3. In this subcontext there are 8 implications, all of them are blue, so all of them have objects in the context which support them. All of these implications follow from the way the scales were defined and all of them have as conclusion the attribute Guitar. The third and the fifth implication also has in the conclusion the attributes Electric Rhythm Guitars, Electric Guitars and Electric Guitars, respectively. This is again obvious because an object with the attribute Heavy Electric Rhythm Guitars obviously has also the attributes Electric Rhythm Guitars and Guitar.

Figure 7 displays a list of association rules with minimal support 0 and confidence 30%.

```
24 < 10 > { } =[90%]=> < 9 > Guitar;
25 < 3 > Electric Guitars Guitar =[67%]=> < 2 > Electric Rhythm Guitars;
26 < 2 > Electric Rhythm Guitars Electric Guitars Guitar =[50%]=> < 1 > Heavy Electric Rhythm Guitars;
27 < 2 > Electric Guitar Solo Guitar =[50%]=> < 1 > Dirty Electric Guitar Solo;
28 < 2 > Electric Guitar Riffs Guitar =[50%]=> < 1 > Dirty Electric Guitar Solo;
29 < 2 > Dirty Electric Guitar Solo Guitar =[50%]=> < 1 > Electric Guitar Solo;
30 < 2 > Dirty Electric Guitar Solo Guitar =[50%]=> < 1 > Electric Guitar Riffs;
31 < 9 > Guitar =[33%]=> < 3 > Electric Guitars;
```

FIGURE 7. List of association rules for the attributes in the Electric instruments subcontext.

Remark 4. The list of association rules contains 8 rules, but the rules from the figure are only part of the list of rules generated by ConExp, which can also be seen by the numbers in front of them: they start from 24, which means there are at least 24 other association rules. A part of the rules not shown here are the implications from the figure above and probably there are also rules with support less than 30%. The first rule says that in 90% of the cases a song has the guitar attribute. This is based on the fact that out of the 10 attributes in this subcontext, 9 are related to guitars and only one to piano.

4.3. **Piano Context.** The third context is the Piano context which contains 6 attributes (Acoustic Rhythm Piano, Acoustic Piano, Subtle Pianos, Electric Pianos, Mellow Piano Timbre, Prominent Rhythm Piano Part) out of which only the Acoustic Piano attribute is multi-valued with values Acoustic Piano and Subtle Acoustic Piano. The scaling was done again using the Ordinal scale, since Subtle Acoustic Piano is still Acoustic Piano. Some of the attributes appeared already in the other two contexts, which is normal, because one attribute can belong to several subcontexts. Similar to the previous two cases, two attributes were also introduced: Acoustic and Electric, to denote the two main categories where a piano can belong to. Still, in this case there are 3 attributes that belong to neither of these categories, the Subtle Pianos, Mellow Piano Timbre and Prominent Rhythm Piano Part does not say what kind the piano is of. Another possibility would have been to include these attributes into both categories. The subcontext contains 124 objects and its concept lattice is presented on Figure 8.

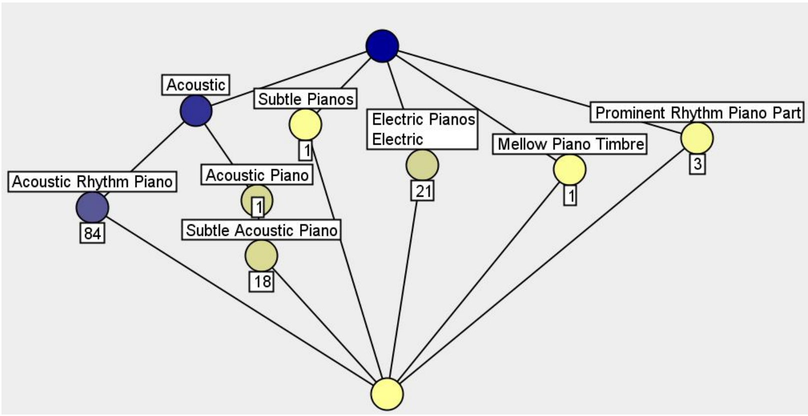

FIGURE 8. Concept lattice for attributes belonging to the Piano subcontext.

5. Towards an FCA Based Music Environment

As it has been stated before, this is a first attempt to investigate the structure of the Music Genome using Conceptual Knowledge Processing methods. Lack of time, the huge dataset and the necessity to optimize the software tool, but also limited space for this presentation, convinced us to restrict ourselves to some relevant topics.

There is more work to be done. We would like to analyze the entire set of attributes, to perform attribute exploration and to compute the stem base over the entire attribute set. Also mining all association rules could display interesting conclusions about how modern music is structured.

Several questions remain unanswered, for example how can we define a similarity measure in order to cluster similar songs?

An FCA based music environment is a research program in this field. The user should be able to choose freely the genomic subset of his convenience, to navigate through conceptual hierarchies, browsing in an intelligent and always conceptual driven way the songs, experiencing music in an entire new way.

References

[1] Carpineto, C., Romano, G.: *Concept Data Analysis, Theory and Applications*, Wiley and Sons, 2004.

[2] Hindemith, Paul, *Traditional Harmony* (second edition), Associated Music Publishers, Inc. New York (1944).
[3] Helmholtz, Hermann L. F., *On the Sensations of Tone*, Fourth (and last) German edition (1877) translated by Alexander J. Ellis as the second English edition (1885), revised, corrected annotated and with additional appendix by the translator, Dover Publications (1954).
[4] Mazzola, Guerino, *Gruppen und Kategorien in der Musik*, Heldermann, Berlin 1985
[5] Mazzola, Guerino, *Geometrie der Toene*, Birkhaeuser, Basel 1990
[6] Mazzola, Guerino, *The Topos of Music-Geometric Logic of Concepts, Theory, and Performance*, Birkhaeuser, Boston-Basel, 1999.
[7] *The Music Genome Project on Pandora:* http:www.pandora.commgp.shtml
[8] Rob Walker, *The Song Decoders*, The New York Times Magazin, October 14, 2009 (available at: http:www.nytimes.com20091018magazine18Pandora-t.html)
[9] Steven Barrie-Anthony (Times Staff Writer), *That Song Sounds Familiar*, Los Angles Times Calendarlive.com, February 3, 2006 (available at: http:msl1.mit.edufurdlogdocslatimes2006-02-03_latimes_pandora_music.pdf)
[10] John Joyce, *Pandora and the Music Genome Project*, Scientific Computing (available at: http:www.scientificcomputing.compandora-and-the-music-genome-project.aspx)
[11] Michael Castelluccio, *The Music Genome Project*, Financial Times, December 2006
[12] List of Alternative Rock Artists: http:en.wikipedia.orgwikiList_of_alternative_rock_artists
[13] ToscanaJ: http:toscanaj.sourceforge.net
[14] Wille, R., *Restructuring lattice theory: an approach based on hierarchies of concepts*. In I. Rival (ed.) *Ordered sets*, Reidel, Dordrecht, Boston 1982, 445–470.
[15] Wille, R.: *Methods of Conceptual Knowledge Processing*, ICFCA 2006, LNAI 3874, Springer, pp. 1-29, 2006.

Faculty of Mathematics and Computer Science, Babes-Bolyai University, Cluj-Napoca, Romania

E-mail address: mzsi0142@scs.ubbcluj.ro

E-mail address: csacarea@math.ubbcluj.ro