International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Aug-24)

http://dx.doi.org/10.12785/ijcds/160152

Applying Process Mining to Generate Business Process Models from Smart Environments

EL KODSSI Iman[1], Sbai Hanae [2] and Kabil mustapha[3]

[1 2 3]*Laboratory of mathematics, computer science and applications (LMCSA), FST Mohammedia, Hassan II university of Casablanca,Morocco.*

Received 26 Nov. 2023, Revised 25 Apr. 2024, Accepted 1 May. 2024, Published 1 Aug. 2024

Abstract: The management of business processes went through several changes. On the one hand, business intelligence (BI) is becoming more popular among businesses as a way to cut costs, boost service quality, and enhance decision-making. On the other hand, we use business process management in a smart environment. So these data sources produced in this environment via sensors, actuators, and other devices are more varied and unstructured, so to apply the process mining techniques, it is necessary to transform them into a structured format. Several works have been done in this direction, and the authors have contributed to the improvement, but the problem is that there is not yet an approach that formalizes the transformation in a general way regardless of the type of sensor data element. Our approach is based on a model-driven architecture (MDA), which allows us to generate source-to-target data transformations. The main objective is to establish the MDA approach via transformation rules based on machine learning techniques.

Keywords: Process mining (PM), raw sensor log, event log, Model Driven architecture (MDA).

1. INTRODUCTION

The emergence of new communication networks and the Internet of Things (IoT) has created a whole new concept in which sensors are linked into our environment, providing researchers and experts with massive volumes of data [1]. In this context, according to the work, the number of objects connected to the Internet of Things, for example, connected cars, industrial machines, meters, consumer products, etc., is expected to grow by an average of 21 % between 2016 and 2022, or 18 billion units globally in the range [2]. This IoT device has profoundly changed personal lives and the world of work. They have introduced advanced sensors, enabling the creation of new models. Moreover, Internet of Things (IoT) devices are capable of generating massive amounts of data.especially as the responsibility for collecting data in a smart environment is difficult and requires careful consideration of the major challenges that emerge in this setting. One of these major issues is managing the vast volume of data created by these intelligent environments, which requires significant expenditure on advanced data analysis technologies and appropriate storage capacity. At the same time, another major obstacle is the variety of technological resources (devices, platforms) and data (formats, sources). It becomes necessary to overcome this variability in order to guarantee optimal use of accessible information and efficient integration. Consequently, the application of creative solutions to these problems is necessary to the success of data collection in an intelligent environment, paving the way for more effective use of data in this dynamic context. It is crucial to implement automated data handling processes to cope with this abundance of information, as it would simply be impossible to manage and analyze such quantities of data manually given the daily time constraints.As mentioned in [3], one specific illustration of a challenge in intelligent settings is the difficulty of raising the standard of services in establishments like hospitals, shopping centers, libraries, and museums. Technology and software-based services are readily available, yet there are still barriers preventing their development in these areas.

In addition, IoT devices need to be connected in a way that suits enterprise operations. If business processes help companies achieve their objectives, IoT capabilities can help organizations establish business processes-aware IoT by linking the physical and digital worlds. However, business processes have evolved due to the explosion of the Internet of Things (IoT) and real-time data collection. But it doesn't stop there. By enabling companies to extract valuable information from this real-time data, process mining has revolutionized process management. Process mining has played a crucial role in the evolution of business process management by bridging the theoretical approaches

E-mail address: elkodssi.iman@gmail.com,sbai.hanae87@gmail.com ,kabilfstm@gmail.com

to process modeling with the complex realities of process execution. Process mining overcomes the conventional drawbacks of manual modeling by automatically creating process models from actual data. through these three layers:

- Process discovery, which produces a business process as an output but requires an event log.
- Process conformance: to determine whether something complies with the provided event log, two entries are needed, such as an event log and the associated business process.
- Process enhancement: this feature makes it possible to add fresh data from event logs to processes.

It's crucial to underline that the effective application of process mining methods requires clearly organized data, in particular an event log. These event logs provide a precise record of every activity and interaction in a specific process. As a result, process mining goes beyond the simple collection of real-time data to provide significant operational insights. Furthermore, to generate business processes via IoT devices, it is crucial to transform sensor logs into event logs to apply PM techniques. Several transformation approaches have been proposed [4], [5], and [6]. Several transformation approaches have been proposed. In this context, we have proposed in [7] and [8] a large comparative study. Among these difficulties is that the study requires detailed browsing behaviors of different target groups. Among these difficulties is that the study needs detailed navigation behavior of different target groups. Moreover, the study of existing approaches also shows that these approaches do not allow the generation of all types of sensor logs. This means that only one type of sensor log can be reviewed. Similarly, another important requirement is to save time and effort and reduce errors by automating the modification of models as much as possible. Our approach is based on MDA, which allows us to generate source-to-target data transformations to make current models more reusable. In this context, model transformation is an essential part of MDA. Thus, in this article, the main objective is to establish the MDA approach, which has the need to formally express the transformations between models, thus making them productive, and to formalize not only the languages in which these are described but also the meta-models describing these languages. Then, using the MDA approach via transformation rules based on machine learning techniques that we have studied in existing approaches, this transition is carried out with a level of Platform-Independent Model to Platform-Specific Model (PIM2PSM) transformations to generate a target meta model, which is an event log. This paper is organized as follows: The first section provides a brief description of the sensors and sensor logs monitored by the MDA approach. Some related works are presented in the second section with a comparative study that summarizes the weak points of the existing approaches. Our approach is described in the third part with a presentation of the proposed meta model. The fourth section shows the applicability of our approach through an illustrative example. The last section concludes the paper and presents future work.

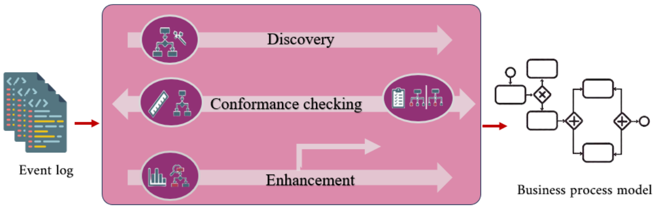


Figure 1. Axes of Process Mining

2. BACKGROUND REVIEW AND FOUNDATIONS

A. Process mining (PM)

Process mining combines data science and process science to improve the study of operational processes using event logs (Figure 1). Process mining begins by collecting data about processes as they occur. Process mining's purpose is to extract an explicit process model from event logs, i.e., the task of creating a process model from a log of events that is compatible with the observed dynamic behavior. Using the event log, various process mining types can be performed in order to generate valuable process-related insights. There are three types of process mining. [9] (figure 1).

The first axe is process discovery. A discovery technique may be used to extract process models that reflect process behavior from an event log [5]. Many algorithms focus on deducing the sequence of process activities from the data. The second type of process mining is conformance [10]. As part of the conformity check, this step seeks to assess the extent to which the data relating to the events corresponds to a given reference model, as well as detect inconsistencies between event log behavior and the process model [9]. The third type of process mining is enhancement. Using process data, these techniques can help enhance and expand an existing process model. Model repair is an upgrade type that enables the alteration of a process model based on event logs.

1) Event log :

The event log standard known as XES is based on XML. For the exchange of events log data between applications, it offers a structured format. Its organization is based on a well-defined flow of activities and events. It contains a log as a collection of activities with the name traces. A trace is a chronological list of events [11]. They are each represented by an XML element. Time stamps, activities, creators, descriptions, and other information are stored in the attributes. For more details, a unique identification for each case named Case id, the activities that each case contained named 'Activity', a reference to when each activity was executed named 'Timestamp'. Aside from this, an event log can additionally include details about the type of event

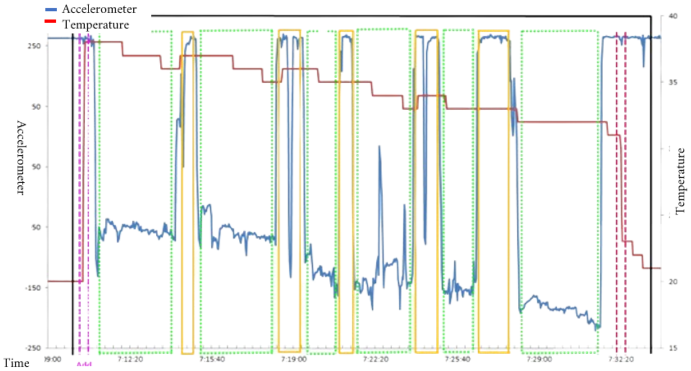


Figure 2. Example of sensor data

named transaction type, the resource connected with the event named resource, and other details about the activity or scenario. [12]

2) Sensor log:

Sensor logs are the output of a device that detects and responds to some type of input from the physical environment. This output can be used to provide information to another system or to guide a process [4]. The events contained in sensor logs are detailed enough, but the actions in mining models should be more abstract. For example, the time series, as a special case of sensor logs, is defined as follows: A data point is a measurement taken by a particular sensor at a given time that records the result of the measurement. Such data can play an important role in improving management.

For instance,[5] provided an illustration of a sensor log that reflects the measurements of three sensors through time in the form of a chronological series. Machine learning (ML) is a technique for teaching computers to process data more efficiently. Machine learning's goal is to learn from data and retrieve useful facts.

3) Machine learning;

To solve data challenges, machine learning employs a variety of algorithms. Data scientists want to emphasize that there is no one-size-fits-all algorithm that is best for solving an issue [13]. The type of algorithm used is determined by the type of problem to be solved.These machine learning algorithms are classified into two main categories: supervised and unsupervised.

Supervised machine learning is the search for algorithms that reason from externally given cases to generate broad hypotheses that subsequently predict future instances. In other words, the purpose of supervised learning is to create a compact model of class label distribution in terms of predictor characteristics. A few samples of supervised learning algorithms are linear regression, logistic regression, artificial neural networks, and support vector machines:

- Linear Regression:In order for the algorithm to anticipate new inputs, linear regression involves fitting a continuous linear function through the data.
- Logistic Regression: Is a known statistical method for determining the contribution of several factors to a pair of outcomes.
- Artificial Neural Networks: To achieve good performance, neural networks use extensive interconnections of ”neurons,” which are basic processing components.
- SVMs (Support Vector Machines): It is used for classification and employs kernel approaches to handle the more challenging scenario of non-linearly separable patterns.
- k Nearest Neighbors:Is one of the machine learning methods that is regarded as being the simplest. The algorithm uses the outputs of its nearest neighbors in the training set to estimate the output of any new input after memorizing the training set.

Unsupervised machine learning: Unlike supervised learning, the algorithms are left to uncover and show the fascinating structure of the data on their own. Unsupervised learning algorithms learn a limited number of features from the data. When new data is introduced, it employs previously learned characteristics to identify the data's class. Unsupervised learning algorithms include, but are not limited to, K-means clustering and dimensionality reduction algorithms.

- K-Means Clustering:This algorithm automatically creates k unique clusters. The variables in the data are grouped together based on relationships between them in this kind of unsupervised learning.
- Dimensionality Reduction Methods:One example is the Principle Component Analysis Algorithm (PCA), whose objective is to minimize the projection error by shortening the distance between each feature and a particular projection line.

4) Model Driven Architecture (MDA) :

A technique for application modeling and generation that has gained a lot of interest recently is called model-driven architecture (MDA). Many organizations are now considering the concepts of MDA [14] and [15], which were promoted by the Object Management Group (OMG), as a way to plan out and manage their application solutions. There are three layers of MDA. Each of the three fundamental layers is essential to the development process. The Computationally Independent Model (CIM) is at the origin of this hierarchy and offers extreme abstraction by capturing essential business elements without worrying about IT considerations. This CIM layer provides a clear understanding of business needs and processes without going into the

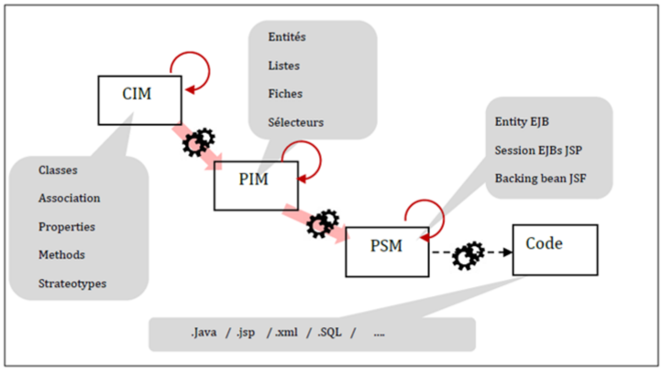


Figure 3. Model driven Architecture

details of technical implementation.

The platform-independent model (PIM) sits on top of the CIM and represents abstract business concepts adapted to a higher level of abstraction while remaining independent of the details of the execution platform. Business logic is encapsulated in the PIM in a generic way.

Finally, the platform-specific model (PSM) is the phase closest to practical implementation. PSM models are adapted to specific platforms by incorporating technological details such as language choice, frameworks, and other platform-specific components. figure 3 shows more details about these layers

This strategy is based on the development of source models and their transformation through various levels of abstraction until the source code is produced automatically. The MDA also includes the definition of several standards [16]. Standardized models and subsequent changes among them are the foundation of model-driven architecture (MDA), which is founded on these ideas. According to MDA, links and rules are represented by a collection of six connected models. It offers a set of guidelines for structuring these model specifications.

- CIM is independent of any computer system. It is the business model or the model of the application domain. The CIM allows the vision of the system in the environment where it will operate, but without entering into the details of the structure of the system or its implementation. It helps to represent what the system will have to do exactly.
- PIM: It is unrelated to any technical platform, such as EJB, CORBA,.NET, and so on, and does not include information about the technologies that will be used to deploy the application. It is a computer model that represents a partial view of a CIM.
- PSM: It is dependent on the technical platform specified by the architect. The PSM primarily serves as a foundation for the generation of executable code for the technical platforms. The PSM describes how the system will use these platforms. There are several levels of PSM. The first, from the transformation of a PIM, is represented by a platform-specific UML schema.

3. STATE OF ART

A. Related works

In the context of business process management, and in particular in an intelligent environment, there are a limited number of works that generate output event logs for any input sensor log. Many methods and approaches have been proposed. for business process management through machine learning techniques to transform inputs into a structured format. [4] [5] [6] [17] In order to automatically discover the process through PM techniques [4], first we will introduce the criteria on which our comparison study is based, and then we will show and debate the comparative study results. The authors of [17] proposed a method for discovering process models from an email log. The method is capable of processing sent and received emails using various process models. The method is able to process emails exchanged in a variety of processing models. Hierarchical clustering and K-means algorithms are used. Emails are grouped based on their subject matter, then based on the process instance they are part of, and finally based on email activity using the hierarchical and k-means clustering algorithms. The study [6] offered a method to cover the creation of an event log for use by process mining tools by extracting information from complicated semi-unstructured databases. The authors of [5], [6], [17], and [18]look at a method for converting sensor data into an event log that can be fed into any process mining algorithm. The mapping of sensor measurements to human activities and the clustering of activities in the same context are the two key issues this study addresses in the application of process mining techniques to sensor data. from a new point. A strategy for discovery processes from raw sensor data was given in [4]. This study proposes a framework for extracting event location sensor data to identify processes and activities. The framework is adaptable enough to be used with any data set, even unprocessed sensor data.

Building on these methodologies. In their own studies, researchers at have highlighted the crucial importance of effective data analysis in an era of ever-increasing amounts of data. In sectors where unstructured data formats are common, this article [19] examines how process mining, machine learning and data mining can be used in concert to extract useful information from unstructured data.

In addition, the researchers make recommendations in their work concerning the main objective of applying process mining (PM) in an intelligent environment[20]. Automatic identification of user behavior patterns from sensor data is the main objective. The main distinction lies in the way sensor logs are converted into event logs in order to use process mining approaches. This process involves adjusting segmentation and granularity. In addition, the integration of process mining (PM) to improve business processes is the main topic of [21]. It draws attention to how contextual elements are overlooked by PM methodologies and seeks to enhance process models with sensor data to improve recommendation making, discovery and compliance verification. The study uses information about routine activities in a smart home to investigate the viability of merging event logs with sensor data. In [22], the concept of improving process models suitable for the development of smart medical devices through the use of process mining is introduced. The method used in [22] enabled the process of developing intelligent medical devices to be optimized by analyzing event logs using process mining techniques.

B. *Criteria for comparison*

The works included in the comparative study are evaluated using seven criteria, which are listed in the following order:

- Types of coverage Domain: determines the method used to prepare event logs before applying process mining techniques. This criterion is used to evaluate the application of machine learning when preparing event logs before applying process mining techniques.
- Axes of PM: It indicates the category of process mining techniques (discovery, conformance, or enhancement) studied in the work. The objective is to identify the most interesting area for research.
- Log Input: Are the data derived from different sources in a smart environment or in another one that is populated with unstructured data?
- Log output: identify the data generated after applying the approaches associated with machine learning algorithms.
- Language or Technologies: indicates the algorithm used to transform raw, unstructured data from the sensor log into event logs with a good level of structuring capable of coping with these traditional PM assumptions

By analyzing these approaches, we find that only one particular category of sensor log is covered by each method. All of these methods use machine learning strategies to support a structured event log, for example, segmentation [5], clustering [17], [16], and classification [24]. For the work [5], the input data is of the time series type. From another point of view, [4] and [1] have treated location data. and there are also approaches that choose to generate event logs via text data [24], [24], [9]. Overall, a common limitation of most work is that the models are not formalized enough to see how their transformation would affect the standard mapping and the domain expert designing the output. Moreover, as it is noted in the input element, each of these approaches has treated the input from a particular point of view. So, our approach proposes first a meta model, which makes it possible to define a sensor log for any category whose objective is to propose an MDA.

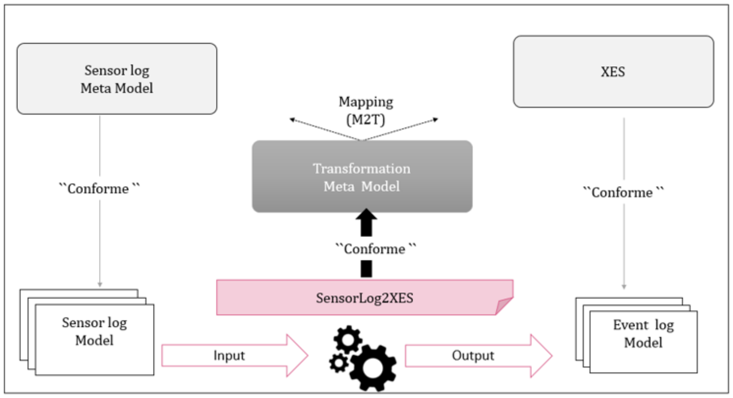


Figure 4. Our Approach for Generating event logs

4. OUR APPROACH: APPLYING PROCESS MINING TO SENSOR DATA IN SMART ENVIRONMENT: MODEL DRIVEN ARCHITECTURE

In order to formalize the mapping to an event log, we propose a framework based on the model-driven architecture. In this approach, we propose a raw sensor log metamodel based on embedded sensors. The code is then generated using M2T (model-to-text) transformations to convert the data and build the meta-model event log that already exists in the literature [25]. Figure 4 highlights the many steps that characterize our method.

The input format for the above table is as follows:

A. *Meta Model Sensor Log*

Given the absence of a meta model associated with a log sensor in the literature, we have proposed a meta model that contains the following elements:

- Time stamp: the time when the event occurred.
- Device: This term refers to the event's resource. This might refer to a physician, a healthcare practitioner, or a medical gadget. The way to obtain the sensor events produced by different sensors becomes more and more complex. These sensors tend to produce massive, distributed, heterogeneous, and complex logs in the new environments of the Internet of Things without a structured schema. This brings an additional level of complexity to what is defined as a sensor raw log, namely: Time series
- Time series: Sensor logs could well be thought of as a time series of sensor readings. A given data point is a measurement collected by a specified sensor at a certain point in time and recorded as the measurement's value.

TABLE I. comparative study

Approach	T.C.D	Axe of PM	Output	Input					Language /Technologie
				L.D	T.S	T.D	Device	Time	
[4]	Transformation	Discovery	E.L	+	-	-	+	+	***
[5]	Transformation	Discovery	E.L	-	+	-	+	+	Centroid , clustering
[6]	Transformation	Discovery	E.L	-	-	+	+	+	Domain-Specific Language
[15]	Transformation	Discovery	E.L	-	-	+	+	+	k-means , language hierarchical
[14]	Transformation	Discovery	E.L	-	-	+	+	+	clustering
[16]	Modelling	Discovery	E.L	-	-	+	-	+	***
[17]	Transformation	Discovery	E.L	-	-	+	-	+	Classification natural language
[18]	Transformation	Discovery	E.L	-	-	+	-	+	Classification
[23]	Transformation	Discovery	E.L	-	-	+	-	+	Classification

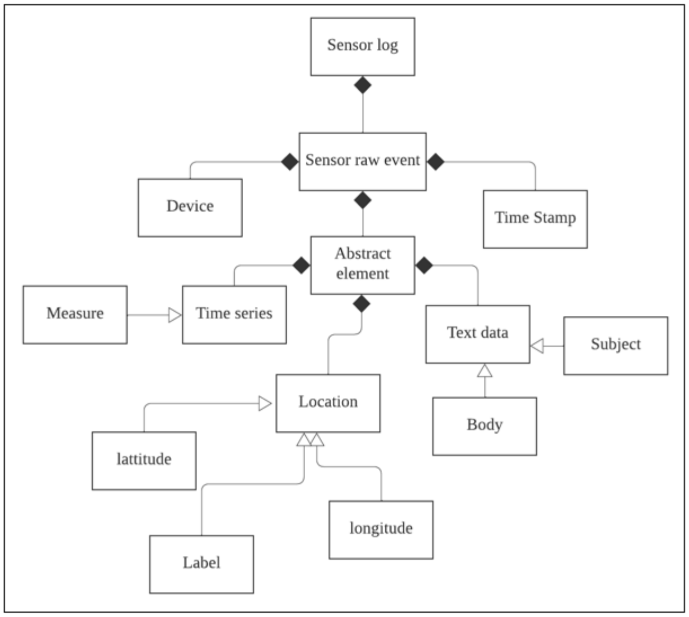


Figure 5. Sensor log meta model

- Text data: this type of data is presented in the form of words, sentences or paragraphs in this type of textual data. This category of data, unlike digital data, offers a semantic richness that allows meanings, contexts and nuances to be conveyed. This textual data can be obtained from a variety of specialized sensors, such as microphones, voice recognition sensors, keyboard sensors or other devices capturing elements of natural language.

- Location: Motion sensors in smart homes and smart industries provide input. This type of event data may or may not include a unique identifier that may be used to determine who caused the event. It may also include a geographical description, such as the locations where the activities were carried out, as well as their frequency and linkages.

B. M2T transformation:

We need to be able to determine the equivalent elements of the source and target meta-models in order to complete this transformation. The mapping rules between sensor raw log and event log meta-model elements are exposed in the following table (see Table 3).

C. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

The first step is to identify the symbols that we will use later, as well as their meaning, namely:

- Idi: The Body of each i email
- Awi: Words associated with an email i
- Si: The subject of each email i m1,...,mn: Measurement set
- Fij: Frequency associated of a i th email and j th word
- C: Model cluster
- D : Geographical distance matrix of all clusters.
- Li : Label associeted of each cluster i
- Ai : Context information of location data .
- Longi/Lati : The weight of word tj in file Idi

The mapping rules are defined in the second phase. These models can be built using the following rules:

Rule 1: AbstractElement2EventLog

TABLE II. Identification Of Corresponding Elements Of Sensor log Meta Model element and event log Meta Model element

Rule	Sensor log Meta Model element	Event log Meta Model element	Transformation Rules name
Rule 1	AbstractElement	EventLog	AbstractElement2EventLog
Rule 2	TextData	Instances	TextData2Instances
Rule 3	Bodysubject	Activity	Bodysubject2Activity
Rule 4	TimeSeries	Instances	TimeSeries2 Instances
Rule 5	DiversityMeasures	Activity	DiversityMeasures2Activity
Rule 6	LocationLog	Instances	LocationData2Instances
Rule 7	PositionGeographiesLabel	Activity	PositionGeographiesLabel2Activity
**	TimeStamp	TimeStamp	***
**	Device	Device	***

The abstract element can be represented in different formats, namely text data, time series, and location, and the list of events in the log is generated by applying three rules: the instances of activities are obtained by applying the rule TextData2 Instances for each TextData element of this composite, and if the abstract element is in time series format, we apply the rule TimeSeries2 Instances. And the last element is applying the rule LocationLog2Instances for each location log element of this composite.

Algorithm 1 AbstractElement2EventLog

```
Input : AbstractElement
This rule creates the event Log which is conform to event log meta model from
A Sensor raw event of sensor log meta model
The name of event is the name of the Sensor raw event
Parameters of a sensor raw event are time resource and abstract element.
If an abstract element is TextData, apply TextData2 instances.
If the abstract element is TimeSerie apply TimeSerie2 instances.
If the abstract element is location log, apply Location-Log2 instances
    Output Event Log
```

Rule 2: TextData2Instances Text data is transformed into an activity instance by applying the ”subjectbody2activity” rule to each body and subject element associated with a text data.

Algorithm 2 TextData2Instances

```
Input : TextData
This rule creates the instances from a text data meta model
The name of event is the name of the Sensor raw event
Parameters of a Sensor raw event are time resource and abstract element
if abstract element is Text Data apply Body&subject2Activity
    Output Instances
```

Rule 3: Body&subject2activities This rule allows the two elements body and subject to be combined to form an activity.

Our proposed approach aims to efficiently clean and organize the data by incorporating transformation rules based on cutting-edge machine learning techniques. Two essential pre-processing techniques are used: tokenization, which divides the text into smaller units like words or phrases, and stemming, which lowers words to their root form for normalization. These first stages make the manipulation of data easier.

The program uses TF-IDF measurement for the extraction of important information. This method reduces the impact of common but uninformative phrases while emphasizing important terms. The goal of using this metric is to highlight how pertinent the data that was taken out is.

Regarding the machine learning algorithm, we have chosen to use a cluster classification strategy. This means that supervised learning techniques in which the algorithm is trained on a collection of labeled data to identify underlying patterns are used to develop transformation rules.

Algorithm 3 Body&subject2activities

```
Input : Body
for each IDi do
    Separate each IDi to words AWi using white space as a delimiter
end for
return tokens AWi
for each AWi in IDi do
    Remove stop words ,numbers , punctuation and lowering capital letters
end for
return vectors
for each awj in IDi do
    Applying numerical Term frequency Inverse Document :
    $Fij \Leftarrow TfIdf(AWi)$
end for
return formulate term document matrix
for each Si attach to IDi do
    for each awj in IDi
    Cutting word SWi by separated words using white space as delimiter process
end for
if SWi similaire a AWi then
    $Fij \Leftarrow 2 * Fij$
end if
Extract the verb-noun pairs are likely to be candidates of being activities
Initialize: each email as an individual clusters $Ci \Leftarrow IDi$
repeat
    /*Compute the proximity matrix */
    Merge(Ci , Cj ) if Ci similar to Cj
    update the proximity matrix
Until only a single cluster remains
return {P C1={ID1,ID2,...,IDk},..., PCn={ IDi,IDj,...,IDm }}
```

```
for each PCi do
    /*Calculate distance between two different emails j and k in the same process model cluster Ci :*/
end for
for each Distance $Eij, Eik$ do
    Applying hierarchical clustering
end for
return { IC1={ID1,ID2,...,IDk}, ...,ICn={ IDi,IDj,...,IDm } }
assumption: an email can contain 0, 1 or more activities.
for each IDi in ICi do Extracted feature of each IDi Applying classification to obtain pertinence sentence
    for each ACi do
        /* choose the top N verb-noun pairs mentioned in the activity cluster*/
        $Activity \Leftarrow verb - noun(IDi)$
    end for
end for
return {ID1={A1,A2,...,Ak},..., IDn={Ai,Aj,...,Am } }
Output instances
```

Rule 5:DiversityMeasures2Activity

The rule ”DiversityMeasures2Activity” allows the creation of an activity from Diversity Measures as described in the following .Time series segmentation is the first step in our job; here, we split the data into discrete sections that are separated by noteworthy occurrences. We can concentrate specifically on the features of each interval thanks to this divide. Subsequently, we incorporate a semantic dimension by labeling these segments in the second phase, according to their temporal features. The final phase involves applying customized algorithms to automatically classify these identified segments. This approach blends classification, labeling, and segmentation.

Algorithm 4 DiversityMeasures2Activity

```
Input : {m1,. . . ,mn}
Create the proposed detailed segmentation of size wi to include the segments ( w1,...,wk ). With k<n.
Each wi corresponds to a set of measures {mi}i<k
for each segment wi do
    Feature selection and calculation of wi
    Characterization of each segment wi
    /*Labeling such that similar segments receive the same label */
    if wi similar to wj thenlabel (wi)=label (wj)
    end if
end for
return set of labelled segments
$Lj \Leftarrow Label(wj)$
Initialisation : Let C1,. . . .Ck be the initial cluster centers
for each point Oi in Ci do
    Calculate the distance between Li and the Oi.
    Assign each Li that is closest to the other centroids Oi.
    $Cnew \Leftarrow newcluster(Li)$
end for
repeat
    Update its center by averaging all of the points oj that have been assigned to it.
    Until convergence
    return {C1, . . . , Ck }
    Identify to each group an activity according to the characteristics of the center of gravity of each cluster:
    $Ai \Leftarrow Activity(Ci)$
        Output Activity
```

Rule 6: LocationLog2 Instances A location log is transformed into an instance, and the list of activities is obtained by applying the rule 'PositionGeographies&Label2activity' for each couple position and label of the composite one .

Algorithm 5 LocationLog2Instances

```
Input : : Location Log
This rule creates the event of event log meta model from a Sensor raw event of sensor log meta model
The name of event is the name of the Sensor raw event
Parameters of a Sensor raw event are time resource and abstract element
if abstract element is LocationLog then
    Apply PositionGeographies&Label2Activity
        Output Instances
```

Rule 7 : PositionGeographies&Label2activity The rule "PositionGeographies&Label2activity" allows the creation of an activity from PositionGeographies and Label .The main guideline that we have incorporated is the use of clustering that is dependent on the geometric location of the data. This method divides similar data into clusters according to how close together they are in geometric space. This rule provides a distinctive viewpoint for finding activity by recognizing spatial links.

Algorithm 6 PositionGeographies&Label2activity

```
Input : locationData
$C \Leftarrow []$
for each locationData ai do
    cnew $\Leftarrow newCluster(lati, longi)$    $addcnewtoC$

    $D \Leftarrow ()$
    for each ci,cj in C: do
        merge clusters that are close to each other and represent the same label
        if label(ci) = label(cj ) then
            $C \Leftarrow C\ ci, cjcicj$
            recompute centroid of the new cluster {ci cj}
            $D \Leftarrow updategeographicaldistancesmatrix$
        end if
    end for
    Identify to each group an activity according to the characteristics of the center of gravity of each cluster
    $Ai \Leftarrow Activity(Ci)$
        Output Instances
```

In order to validate our approach, we present a case study for implementing the proposed automated generation in the next section.

5. METHODOLOGY

Our methodical approach to autonomous process discovery begins with a thorough study of the data in an intelligent environment. This section outlines the main steps of our technique and emphasizes how this article's rules enable automated business process discovery. These phases are visually illustrated in Figure, which also offers a visual summary of our process discovery methodology. Automated process discovery is the foundation of our procedure. Using the methodology, we have presented, we carry out an in-depth study of the data accessible in an intelligent environment during the first phase, focusing particularly on mapping during the second stage. This preparatory stage is essential to guarantee the quality of the process mining results in the next stage (stage 3), which aims to extract behavioral patterns from the discovered event log. We ensure that a well-defined business process emerges from the intelligent analysis of data in an initial mapping context based on well-defined rules by following this logical sequence using our approach.

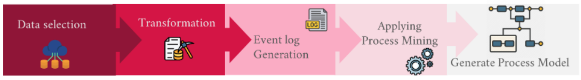


Figure 6. methodology of our approach .

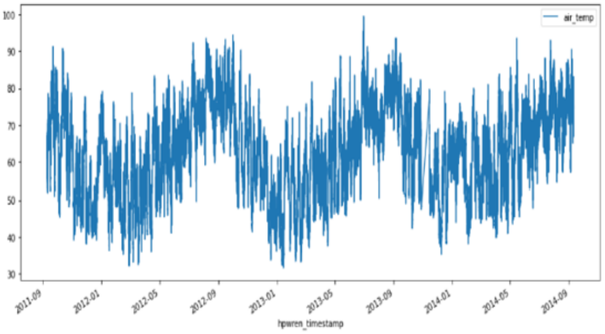


Figure 7. Temporal Variation of Temperature Measurements from Sensor Data in the Context of Climate Adaptation

6. EVALUATION

For our case study, we use one of the elements of the sensor log for sensors that measure the temperature in the context of climate adaptation in different cities. For the 16 automatic adaptations of the movement of the intelligent windows, the latter makes it possible to react according to the environmental conditions surrounding the installation. Our objective is to apply the proposed rules for generating a model of an XES event log. At the CIM level, one of the input elements of the sensor log is a time series. This type contains:

- Case ID: a unique number for each line.
- Date and Time: timestamp of the measurement.
- Temperature: temperature measurement at the time stamp.
- Humidite (%): air humidity measurement at the time stamp

Time series are divided into separate windows to transform sensor data into a set of measurements. These small windows are then labeled. To determine which actions were performed during these observations, the data segments collected in the sensor log are grouped together on the basis of the properties of each segment. To determine which actions were performed during these observations, the data segments collected in the sensor log are grouped together on the basis of each segment's properties. And for each class, we assign the associated activity. We used our recommendations to extract the event log. At the CIM layer, we get an event log that conforms to the XES.

As perspective and to illustrate the modeling example, figure14 shows the modeling process using BPMN2.0.

7. Disscusion

Process mining techniques are applied to gain a deeper understanding of resource, IoT and data processes. This uncovers optimization opportunities and the changes needed to make effective use of these strategic aspects.

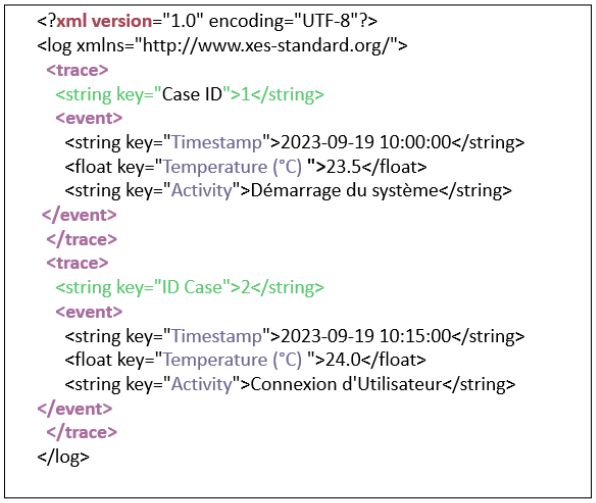

```
<?xml version="1.0" encoding="UTF-8"?>
<log xmlns="http://www.xes-standard.org/">
 <trace>
  <string key="Case ID">1</string>
  <event>
   <string key="Timestamp">2023-09-19 10:00:00</string>
   <float key="Temperature (°C) ">23.5</float>
   <string key="Activity">Démarrage du système</string>
 </event>
 </trace>
 <trace>
  <string key="ID Case">2</string>
  <event>
   <string key="Timestamp">2023-09-19 10:15:00</string>
   <float key="Temperature (°C) ">24.0</float>
   <string key="Activity">Connexion d'Utilisateur</string>
</event>
 </trace>
</log>
```

Figure 8. Sample event output showing temperature Measurements from Sensor Data in the Context of Climate Adaptation to XES format

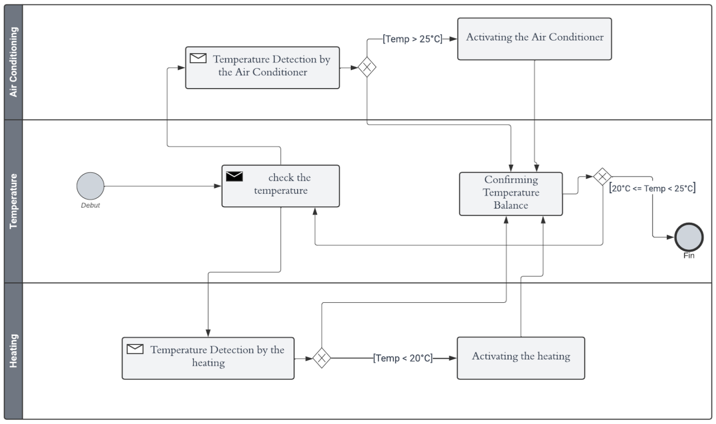

Air Conditioning
Temperature
Heating
Temperature Detection by the Air Conditioner
[Temp > 25°C]
Activating the Air Conditioner
Debut
check the temperature
Confirming Temperature Balance
[20°C <= Temp < 25°C]
Fin
Temperature Detection by the heating
[Temp < 20°C]
Activating the heating

Figure 9. The process model for an temperature Measurements system

We decided to use process mining techniques on unstructured data, namely sensor logs, as part of our strategy. These logs, which are produced by sensors in an intelligent environment, often include a significant amount of unstructured data on measurements, events and environmental circumstances.

Converting unstructured data into a structured event log is a crucial step in our process. The aim of this transformation is to organize sensor log data in such a way that process mining tools can use it to its full potential. To achieve this, the data frequently needs to be normalized, organized and enhanced to provide a logical, chronological flow of events.

Using modern machine learning techniques, our approach aims to fundamentally revolutionize business process management in an intelligent environment. Its main objectives are to correctly generate business processes and organize data, as well as to proactively discover inefficiencies and foster organizational agility. Here are some of our unique qualities: The accuracy of our model is maintained by carefully converting unstructured data into an event log, ensuring accurate assessments for process management in a perceptive and dynamic environment.

In contrast to , our approach stands out for its enhanced scalability functionality, enabling registered users to process sensor data in intelligent environments. This special feature enhances the adaptability and flexibility of our solution, enabling more in-depth modification of the process to suit specific business requirements. Our method is unique in that it can convert any type of sensor into structured data, unlike other systems which often focus on processing a single sensor element. As a result, our solution is highly efficient and ubiquitous, making it a creative and adaptable choice for intelligent processing of sensor data in dynamic, intelligent environments.

In summary, our solution offers significant advantages over other tools due to its sophisticated scalability functionality and flexibility in process modeling, as shown by the comparison with other tools. These significant improvements distinguish our solution from competing platforms, and make it a more flexible and customized choice for enterprise process management.

The main objective of our current strategy is to map data to locate an event log so that Process Mining can be used to create a business process. It is imperative to draw attention to an inherent weakness of this methodology, namely its over-emphasis on data transformation, to the detriment of full utilization of the intelligent environment. In order to better organize and contextualize the extracted data, we intend to include an ontology in our output model as part of future advances. In addition, we intend to study the feasibility of augmenting our model by including Internet of Things (IoT) components, in order to provide a more comprehensive and astute viewpoint in business process modeling.

8. conclusion and Future Work

This model represents an indoor temperature management process based on measured values. It uses an exclusive gateway to make decisions based on the detected temperature range and act accordingly, activating either heating or cooling or doing nothing if the temperature is already within a comfortable range. The process ends once one of these actions has been taken, confirming that the temperature is balanced.

With the evolution of the Internet of Things and the strong presence of sensors in various environments, we are encouraged to integrate product data for the automatic discovery of business processes and to exploit the process information by providing a better understanding via the use of process mining. However, because the data in our resource becomes obviously voluminous and unstructured, it is not compatible with process mining. To solve this problem, we propose an MDA approach for the generation of an event log from the sensor log, whose objective is to automatically discover a business process. We proposed a set of mapping rules for applying the M2T transformation of the sensor log model into the event log model. Nevertheless, our model has certain limitations, such as the fact that it does not take into account Internet of Things-related elements in the results obtained. as a perspective, One option for improvement would be to enrich the events captured, enabling the creation of a conscious IoT process. This potential development could enhance the model's adaptability to the specificity of IoT environments, improving the relevance of its analyses and process management in these dynamic contexts. Our future work will give a semantic representation to describe the concepts related to the IoT and the elements of an executable business process described in BPMN. We use standard semantic technologies, in particular anthologies.

References

[1] O. Vermesan and P. Friess, Eds., *Internet of Things Applications - From Research and Innovation to Market Deployment*. Taylor & Francis, 2014, accepted: 2022-11-28T16:03:08Z.

[2] B. Gupta and M. Quamara, "An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 21, p. e4946, 2020, number: 21 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.4946. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.4946

[3] E. Faris Yahya and M. Mohammed, "The Challenges of Smart Environments and Future Trends Supported by the Internet of Things," May 2023.

[4] A. Koschmider, D. Janssen, and F. Mannhardt, "Framework for Process Discovery from Sensor Data."

[5] M. L. van Eck, N. Sidorova, and W. M. P. van der Aalst, "Enabling process mining on sensor data from smart products," in *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*, Jun. 2016, pp. 1–12, iSSN: 2151-1357.

[6] Valencia-Parra, B. Ramos-Gutiérrez, J. Varela-Vaca, M. T. Gómez-López, and A. G. Bernal, "Enabling Process Mining in Aircraft Manufactures: Extracting Event Logs and Discovering Processes from Complex Data."

[7] E. Iman, M. D. Laanaoui, and H. Sbai, "Applying Process Mining to Sensor Data in Smart Environment: A Comparative Study," in *Innovations in Smart Cities Applications Volume 6*, ser. Lecture Notes in Networks and Systems, M. Ben Ahmed, A. A. Boudhir, D. Santos, R. Dionisio, and N. Benaya, Eds. Cham: Springer International Publishing, 2023, pp. 511–522.

[8] I. Elkodssi, M. D. Laanaoui, and H. Sbai, "Toward a New Framework for Process-Aware IoT Discovery," in *Innovations in Smart Cities Applications Volume 6*, ser. Lecture Notes in Networks and Systems, M. Ben Ahmed, A. A. Boudhir, D. Santos, R. Dionisio, and N. Benaya, Eds. Cham: Springer International Publishing, 2023, pp. 793–803.

[9] M. de Leoni, W. M. P. van der Aalst, and M. Dees, "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs," *Information Systems*, vol. 56, pp. 235–257, Mar. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437915001313

[10] J. Wang, R. K. Wong, J. Ding, Q. Guo, and L. Wen, "Efficient Selection of Process Mining Algorithms," *IEEE Transactions on Services Computing*, vol. 6, no. 4, pp. 484–496, Oct. 2013, number: 4 Conference Name: IEEE Transactions on Services Computing.

[11] N. Martin, M. Swennen, B. Depaire, M. Jans, A. Caris, and K. Vanhoof, "Batch Processing: Definition and Event Log Identification."

[12] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *Journal of Biomedical Informatics*, vol. 61, pp. 224–236, Jun. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046416300296

[13] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017, number: 4 Conference Name: IEEE Communications Surveys & Tutorials.

[14] A. Karkouch, H. Mousannif, H. A. Moatassime, and T. Noel, "A model-driven architecture-based data quality management framework for the internet of Things," in *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*, May 2016, pp. 252–259.

[15] C. Dickerson and D. N. Mavris, *Architecture and Principles of Systems Engineering*. CRC Press, Apr. 2016.

[16] Z. Valero-Ramon, C. Fernandez-Llatas, B. Valdivieso, and V. Traver, "Dynamic Models Supporting Personalised Chronic Disease Management through Healthcare Sensors with Interactive Process Mining," *Sensors*, vol. 20, no. 18, p. 5330, Jan. 2020, number: 18 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/20/18/5330

[17] D. Jlailaty, D. Grigori, and K. Belhajjame, "A framework for mining process models from emails logs," Sep. 2016, issue: arXiv:1609.06127 arXiv:1609.06127 [cs]. [Online]. Available: http://arxiv.org/abs/1609.06127

[18] M. ELLEUCH, N. ASSY, N. LAGA, W. GAALOUL, O. ALAOUI ISMAILI, and B. Benatallah, "A Meta Model for Mining Processes from Email Data," in *2020 IEEE International Conference on Services Computing (SCC)*, Nov. 2020, pp. 152–161, iSSN: 2474-2473.

[19] A. Koschmider, M. Aleknonytė-Resch, F. Fonger, C. Imenkamp, A. Lepsien, K. Apaydin, M. Harms, D. Janssen, D. Langhammer, T. Ziolkowski, and Y. Zisgen, "Process Mining for Unstructured Data: Challenges and Research Directions," Nov. 2023, arXiv:2401.13677 [cs]. [Online]. Available: http://arxiv.org/abs/2401.13677

[20] Y. Bertrand, B. Van den Abbeele, S. Veneruso, F. Leotta, M. Mecella, and E. Serral, "A Survey on the Application of Process Mining to Smart Spaces Data," in *Process Mining Workshops*, ser. Lecture Notes in Business Information Processing, M. Montali, A. Senderovich, and M. Weidlich, Eds. Cham: Springer Nature Switzerland, 2023, pp. 57–70.

[21] R. Elali, E. Kornyshova, R. Deneckère, and C. Salinesi, "Mining Contextual Process Models Using Sensors Data: A Case of Daily Activities in Smart Home," in *Information Systems*, ser. Lecture Notes in Business Information Processing, M. Papadaki, P. Rupino da Cunha, M. Themistocleous, and K. Christodoulou, Eds. Cham: Springer Nature Switzerland, 2023, pp. 409–425.

[22] A. Schidek and H. Timinger, "Concept for Process Mining for smart Medical Device Development," in *2022 IEEE 7th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*, Aug. 2022, pp. 161–167, iSSN: 2687-6817.

[23] F. Fonger, "Process Mining on Distributed Time-Series Data (PhD Proposal)."

[24] Y. Rashnavadi, S. Behzadifard, R. Farzadnia, and S. Zamani, "Discovering Business Processes from Email Logs using fastText and Process Mining," May 2020. [Online]. Available: https://www.techrxiv.org/articles/preprint/Discovering_Business_Processes_from_Email_Logs_using_fastText_and_Process_Mining/12283835/1

[25] P. Lam, H. Ahn, M. Park, K.-S. Kim, and K. Kim, *A Proportional Process Mining System*, Jul. 2019.

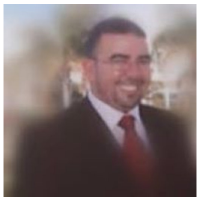
Author 1 EL KODSSI Iman Iman, E. K., Hanae, S. (2023, May). A study of extending BPMN for IoT-aware process modeling. In Proceedings of the 6th International Conference on Networking, Intelligent Systems Security (pp. 1-4).

Iman, E., Laanaoui, M. D., Sbai, H. (2022, October). Applying Process Mining to Sensor Data in Smart Environment: A Comparative Study. In The Proceedings of the International Conference on Smart City Applications (pp. 511-522). Cham: Springer International Publishing.

Elkodssi, I., Sbai, H. (2023). Toward a new semantic framework for internet of things-aware business process discovery. In ITM Web of Conferences (Vol. 52, p. 02001). EDP Sciences.

Author 2 SBAI Hanae Sikal, R., Sbai, H., Kjiri, L. (2018, October). Configurable process mining: variability Discovery Approach. In 2018 IEEE 5th international congress on information science and technology (CiSt) (pp. 137-142). IEEE.

Sbai, H., Fredj, M., Kjiri, L. (2014). A pattern based methodology for evolution management in business process reuse. arXiv preprint arXiv:1403.6305



Author 3 Kabil Mustapha Leroux, J., Schmitz, S. (2015, July). Demystifying reachability in vector addition systems. In 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science (pp. 56-67). IEEE. Kabil, M. (1992). Enveloppe injective de graphes et de systèmes de transitions et idéaux de mots (Doctoral dissertation, Lyon 1