ESCUELA POLITÉCNICA NACIONAL

THESIS

For the award of the degree of

DOCTOR OF PHILOSOPHY IN INFORMATICS

Resolution RPC-SO-43-No.501-2014 of the Consejo de Educación Superior

Presented by

OSWALDO EFRAIN DÍAZ RODRIGUEZ

Thesis supervised by **María Gabriela Pérez Hernández** Professor of the Escuela Politécnica Nacional

Intention Mining of an Information Systems User

Oral examination by the following committee:

Edison Fernando Loza Aguirre, Ph.D.
Escuela Politécnica Nacional (EPN)

Diana Cecilia Yacchirema Vargas, Ph.D.
Escuela Politécnica Nacional (EPN)

Antonio Alexander Franco Crespo, Ph.D.
Escuela Politécnica Nacional (EPN)

Mónica Katiuska Villavicencio Cabezas, Ph.D.
Escuela Politécnica del Litoral (ESPOL)

Juan Alfonso Lara Torralbo, Ph.D.
Universidad de Córdoba (UCO)

DECLARATION

I hereby declare under oath that I am the author of this work, which has not previously been presented for obtaining any academic degree or professional qualification. I also declare that I have consulted the bibliographic references included in this document.

Through this declaration, I transfer my intellectual property rights corresponding to this thesis, to the Escuela Politécnica Nacional, as established by the Intellectual Property Law of Ecuador, its Regulations and the current institutional norms.

Oswaldo Efraín Díaz Rodríguez

CERTIFICATION

I certify that Oswaldo Efraín Díaz Rodríguez has carried out his research under my supervision. To the best of my knowledge, the contributions of this work are novel.

María Gabriela Pérez Hernández

ADVISOR

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

GLOSSARY

Business in general: We considered the business activities of all businesses and structured the knowledge base (KB) proposed in [1], which contains activities of all kinds. The KB was used to validate the activities that take place in a specific sales business through the proposed technique. In this case, a sales business was chosen, but any other business can be used as long as it has the necessary data for modeling.

Business rules: The formal declarations of a business, which allow commercial objectives to be achieved through developed activities [2]. Based on the business rules, the heuristic rules that validate user activities are deduced.

Current Process Model: A process model defined by business staff, which is not set in stone. On the contrary, it can and should be updated periodically, according to the resilience level of a company [3].

Current processes: (Business process manual) the promulgated processes that are used to develop the activities and reach business objectives [4].

Database: According to relational design, non-redundant and interrelated data sets are stored in predefined structures and processable by several applications through integrations and safeguards [5]. For dimensional design, a database is a set of facts, dimensions and measures structured in stars, constellations and snowflakes that can store large amounts of data for Business Intelligence processing [6].

Event log: This file contains a chronological record of daily business activities. For objects of the present study, it is the event log related to the activities of a generic sales process [7], [8], [9].

Event: The fundamental entity of observed physical reality represented by a point designated by three coordinates of place and one coordinate of time in the space–time continuum postulated by the theory of relativity; a subset of the possible outcomes of an experiment [10].

Generic Event Log (GEL): In a similar way to the Generic Windows Event Log [11], the concept and use of the “Generic Event Log” has spread in information systems, and it could be indicated that GEL is an event log generated by default by the information system in any business.

Ideal Process Model: In this theoretical model, business rules are fully complied with in an ideal environment, where entropy is reduced to zero and activities are carried out strictly according to the norm (legacy process) [12].

Information System: A set of applications developed or adapted to suit business policies and rules [13], which the user uses to carry out their daily activities in the business.

Intention Mining: Technique to identify user strategies and formalize their intentions [14], [15], in the execution of business activities.

Process instance: This is a subset of process activities where the presence of the process start and process end activities is mandatory [16].

Process Mining: This is defined as the possibility of discovering actual processes, verifying their conformity with current processes and improving these processes using process mining tools [7] and event logs.

Process Traces: A trace of a process corresponds to an executed instance of the process; it is the sequence of activities that are executed to perform a process in a specific instance to satisfy certain requirements and for a specific user [16].

Quality Event Log (QEL): The QEL is corrected and debugged for a specific purpose and to improve the quality of process mining and intention mining [17].

Real Process Model: A process model obtained from the event log [15], [18]. In this log, the activities that are actually carried out in business are recorded, including the activities according to the current processes plus the user strategies.

Real Process: Set of activities carried out on a daily basis by the user, based on the current process [19] and using their own strategies to achieve their goals in accordance with business objectives.

Specific Business: A formal business, where daily activities are developed using an information system and some databases. As a result, the processing of activities in a business and the generic event log (GEL) [8], [20] are produced, where business activities are stored.

User intentions: Without losing track of business rules and according to user task maturity and acquired expertise, users perform their daily activities with the intention of reaching their goals according to business objectives, allowing saving resources and improving the quality of results. Regarding the execution of a certain process, as well as daily activities, the user utilizes a set of strategies that allow them to fulfill their intentions [15]. Each user intention can be fulfilled with several strategies, and each strategy can be used to fulfill several intentions.

User strategies: At the beginning, new users develop their activities according to current processes, but as users gain experience, they try to perform the same activities using a better method; thus it helps them save resources, specifically, the execution time of their daily tasks. This is achieved through the use of their own strategies [21], [22], [23] such as end user tools, pre-processed formats and forms (templates), or pre-processed intermediate data results. Consequently, the event log keeps traces that do not agree with current business process models.

Users: Business information system users carry out daily activities for specific businesses through an information system [2], using both current business processes and their own strategies.

RESUMEN

En un escenario (organización, empresa, negocio, escuela, etc.), donde el principal protagonista es el usuario, se abordó el nuevo tema de la "minería de intenciones" en el contexto de los sistemas de información empresarial. Los datos que se procesan en el desarrollo de las actividades empresariales diarias permiten generar un registro de eventos (transacciones). Este registro de eventos contiene las actividades realizadas de acuerdo con el manual de procedimientos empresariales, así como las estrategias que los usuarios proactivos utilizan para mejorar su rendimiento, de acuerdo con sus perfiles, políticas y reglas empresariales.

Se desarrollan métodos de revisión, diseño y minería (procesos, sentencias y estrategias) para inferir las intenciones del usuario del sistema de información empresarial. Se crea una base de conocimientos para cualquier negocio en general. A partir del registro de eventos de ventas (negocio específico), se extraen las estrategias de un usuario. Las estrategias del usuario se verifican, validan y ponderan. Finalmente, a partir de estas estrategias del usuario, se infieren las intenciones del usuario del sistema de información comercial de ventas

ABSTRACT

In a scenario (organization, company, business, school, etc.), where the main protagonist is the user, the new topic of "mining of intentions" was addressed in the context of business information systems. The data that are processed in the development of daily business activities allow a log of events to be generated (transactions). This event log contains activities carried out in accordance with the business procedures manual as well as strategies that proactive users use to improve their performance, according to their profiles, policies and business rules.

Review, design and mining methods (processes, sentences and strategies) are developed to infer the intentions of the business information system user. A knowledge base is created for any business in general. From the log of sales events (specific business), a user strategies are extracted. The user strategies are verified, validated and weighed. Finally, from these user strategies, the user intentions are inferred from the sales business information system.

Keywords: information systems; business processes; event logs; user strategies; intention mining; business rules

PROLOGUE

In development of human activities (user activities), reaching the objectives is mandatory. On some occasions, there is an intention to save resources, show efficiency/effectiveness, or achieve an objective using a different method than the one established, among others. In the field of business, specifically in the development of activities through information systems, a record of activities is generated; commonly called event log. The users execute the activities observing the business rules, but they also use their own strategies to achieve their objectives according to their functions (user profiles). Therefore, in the event log, the activities that are executed in the business (activities according to the business rules and user strategies, among others) are saved. From the event log, the strategies of the users are extracted to infer their intentions; Based on the intentions of users, it is possible to infer their behavior, identify new information requirements, and develop process-flexibility business information systems. The main contributions of this work can be summarized as follows: A method of conducting a literature review of intention mining in the information systems field; methodology for designing relational databases, indispensable for the generation of a consistent generic event log; a method of generating a quality event log from a generic event log; a knowledge base building method, developed via the sentence mining with on natural language processing tools and an intention mining method from the information system user strategies.

CHAPTER 1

1. INTRODUCTION

Section 1.1, entitled “Context and Motivation”, provides an overview of information systems in the business setting. Section 1.2 identifies the lack of flexibility in these information systems and their problems. Section 1.3 describes these business information systems, which are centered on the user. In Section 1.4, a business scenario is addressed through a theoretical and conceptual framework. In Section 1.5, the research proposal on the topic "mining of intentions" is presented: the general objective, specific objectives, research methodology and relevance of the research topic are presented. In Section 1.6, the research methodology is specified. Section 1.7 details contributions from this field of research. Finally, the thesis is outlined in Section 1.8.

1.1. Context and Motivation

Modern companies, in order to develop their daily activities, need information systems, whose fundamental architecture must involve at least the following elements. The business domain defines the business environment for a sales business, particularly one where the sales process is the object of study and applications. In a business infrastructure that allows data processing and the generation of information in accordance with the business requirements, the most important elements for the purpose of this study are databases and information systems. Therefore, database managers facilitate access to their stored data in order to process these data in information systems and meet the information requirements of the business, among other things. However, the most important aspect of the present study is the event log, which information systems either generate automatically or through a previous configuration established for this purpose [24].

In this specialist field, according to the business abstraction levels [25], [26], a general business scenario is formulated. However, inside this scenario, the scope of a specific and particular business is limited. In this study, a sales business is addressed.

The focus and scope of the business scenario are defined [27] for one of the most typical types of modern business, the sales business, whose processes require technological support in order to develop activities through an information system. Scenario policies define the business approach to general sales, and business rules define the scope and domain of the business, whose business environment is defined by the type of product or service that is sold.

Data, in different forms, formats, models and presentations, are the input of the information systems, which in order to be processed and meet the information needs of users and business rules, need to be stored in predefined structures according to the design of the relational and dimensional databases. In one study [5] a methodology was established for the relational design of databases in the normal third form [28] and for business environments. For a dimensional design [6], one of the most important and relevant methodologies was specified by Kimball, who pioneered this type of dimensional database design, which has been the basis for the implementation of the data warehouse.

Through this information, as a product of the execution of processes through the information system, users have the information to fully comply with daily activities, which allow them to meet the business rules, according to the scope and responsibility of their functions (profiles).

Some users develop their activities using information systems based on the current business processes (procedure manuals), and these data are available in repositories and databases installed and configured in the business environment.

1.2. Problem Statement

In the business world, information system users try to develop their activities in the shortest time possible. Each user does their job, developing activities according to their profile and business rules [15], [23]. However, information systems do not provide adequate facilities [12], [13], so users must use their own strategies to achieve their work goals.

The user executes the specified activities in a manual of company procedures [19], but they also use their own strategies; these strategies underlie user intentions [9]. For business administrators, it is necessary and important to identify user intentions and understand their behavior for monitoring and control purposes.

For this reason, the present study aims to identify and formalize user intentions based on their strategies extracted from the event log [7], which is generated by the business information system.

1.3. Background

The field of information sciences studies information processing practices and defines theories of the knowledge of the universe of data to understand (content, forms, nature, meaning and essence), characterize, and process these data, as well as generating information and knowledge of all information technologies (information systems). While information sciences are responsible for describing the universe of data and information and using information technologies to trial and test their theories, information technologies modify the universe (data and information) and adapt it to the needs of humanity. To satisfy the information requirements in a given business, users develop their daily activities using information systems, and according to their profiles, execute the necessary processes to achieve their goals, which are framed in the company's objectives and strictly aligned with business policies and rules. The processes defined and structured based on these business rules and legacy processes (if applicable) are the so-called current processes [4]. Based on these, the daily activities in the company are developed. However, users with the intention of saving resources, especially time, develop activities based on their current processes and strategies, which gives rise to the real process, whose traces are stored in the event log. Using process mining techniques [7], [19] from the event log, the actual processes can be discovered, based on which recommendations can be issued, and current process can be monitored and improved. On the other hand, using traditional mining techniques, based on the same event log but refined with quality attributes, user strategies can be discovered and formalized. Based on user strategies, user intentions can be inferred.

1.4. Theoretical and Conceptual Framework

An organization, in general terms, can be any set of resources (technological platform, users, physical infrastructure, etc.) organized to meet some objective. It can be governmental, private, mixed, and with or without profit. An enterprise can be defined as a for-profit organization, where a statute defines its field of action and business guidelines [29]. A business can be considered a company or a part of it [30], according to the following considerations. A process is strictly defined by enterprise policies and rules that govern a business. A business is a set of specific activities (current process) that define how to develop activities (know-how), and a strictly well-defined business (current process model) may correspond to an enterprise and vice versa. For a multi-business enterprise, each business has its own current process model; consequently, in development of this research topic, the term "business" will be used (to refer enterprise). Additionally, for reasons of affinity and compatibility with the current literature, the research topic addresses the business process model and the business term "sell". In the glossary, the concepts that are used in this study are defined.

1.5. Research Proposal

1.5.1 General Research Aim

The aim of this study is to propose a method of mining the user strategies of business information systems and to infer their intentions.

1.5.2 Specific Research Aims

- Characterize the current process for a specific business
- Define the real process of the specific business from an event log
- Identify user strategies
- Build a general knowledge base for any business
- Verify user strategies according to the knowledge base
- Validate and weighting; the user strategies, according to the heuristic rules of the specific business
- Inferring user intentions.

1.5.3 Research Questions and Hypothesis

In accordance with the paradigms of the qualitative research proposed by [31]. The research questions (RQ) are defined in an ontological (RQ1), epistemological (RQ2) and methodological (RQ3) way, using the current process, actual process, and user strategies of a business information system the as research objects. In addition, in order to meet the objectives, hypotheses (H) are defined to answer each research question.

RQ1. In a business, are the current process, the actual process, and the functions of the person who executes them known and promulgated?

H1.1. Through policies and business rules, it is possible to know the current business process [32].

H1.2. The business information system processes the database and generates an event log by default [5], [20].

H1.3. Through process mining applied to the event log, it is possible to obtain the real business process model [15], [33], [34].

H1.4. From differences between the current process and real process of the business, user strategies are defined [19], [20], [34].

RQ2. In a business information system, is there a relationship between the current process, the real process and the users who execute them?

H2.1. Using process mining techniques [35], it is feasible to establish the difference between the current business process model and the real business process model [15].

H2.2. Based on the differences the business processes model (current and real), it is feasible to identify the user strategies [34].

RQ3. How to infer user intentions? [34]

H3.1. It is possible to develop a method to structure a general knowledge base for any business [36].

H3.2. It is possible to extract the user strategies from quality event log of the specific business [20].

H3.3. User strategies can be validated and weighted through the heuristic rules of the business [37].

H3.4. From the weighted strategies of the user, it is possible to infer their intentions [34].

1.6. Research Methodology

The novel research topic of intention mining [19] is the main focus of this study; this methodology is proposed based on process mining, business knowledge base, event log and the formalization of user strategies (Figure 1-1). According to the design science research methodology [38], [39], identify the problem, define the aims, and develop and validate a method . Below, the principal activities are specified:

- Research introduction

- Defining the research topic and its feasibility
- Method of literature review and state of the art
- Method of a quality event log for intention mining
- Proposed method of intention mining
- Proposed method validation.

1.6.1 Relevance of Approach

Today, the development of information systems involves a user, depending on their profile, and procures the flexibility of processes through feedback [40] and the alignment of the development of activities with business objectives. Identifying user behavior can support the decision making of managers, administrators, developers and users.

- **Requirements Analysis**

Identifying user behavior can allow new ways of executing business activities to be established in order to improve and innovate traditional processes. The current process model corresponds to the formalized process of a business prior to the development of a new or redesigned information system. For the option of a new system, a company carries out a process survey to identify the current situation of a business or businesses (the company's purpose) in order to set a baseline for the development of a new information system [41]. In the case of a redesign, in addition to the current business process, user strategies are considered as requirements for the design and implementation of the new version of an information system.

- **Process Design**

Based on the actual process model, the implementation of improvements or redesign of the current process is possible. Nowadays, the monitoring of the development of activities that are executed through the information system is frequently carried out through the processing of an event log [42], which is currently used to carry out business audits. In this study, the event log is used to extract user strategies, which is used in the formalization of “user’ intentions”. These user intentions can improve the current process model or identify the need for a redesign of this process. Following the development of a new version of the information system, the company would no longer have the need to carry out a new survey of the current process. Based on the user intentions, a new design could be obtained to satisfy current information requirements and new ways of generating information.

- **Process Management**

Businesses must generate revenue by saving resources and reducing execution time and process costs. Historically, in information systems, the development of each activity has its own cost [7]. Consequently, the cost of executing business processes is already defined. However, a user can develop this process by replacing a set of activities with one of its strategies (intentions), which makes the cost of executing the process different (lower or higher) from the pre-established cost. The same would happen with the execution time of the process, although in this case, there is no pre-established time for the execution of the process. This corresponds to the administrators (middle managers) of the process, who monitor processing costs versus execution times and maintain a balance of satisfaction between business customers and information system users.

- **Application**

The supervision of user activities and suggestions and recommendations based on the history of their performance are recorded in the event log [15]. Intent mining can determine user strategies and especially the intentions of these applications through the execution of the daily business process activities. From the analysis of the user intentions, user behavior with respect to these business process could be determined. Consequently, it would be possible for business administrators to make suggestions and recommendations regarding decision-making and measures of management level control [43].

1.6.2 Expected Results

The method of mining user intentions of the business information system is used, and this method is developed and validated.

1.7. Research Contributions

By fulfilling the proposed objectives and responding to the research questions by satisfying our hypotheses, these are the contributions of the present study:

- A method of conducting a literature review of intention mining in the information systems field.

- Methodology for designing relational databases, indispensable for the generation of a consistent GEL;

- A method of generating a QEL from GEL;

- A knowledge base building method, developed via the sentence mining with on NLP tools;
- An intention mining method from the information system user strategies.

1.8. Thesis Outline

According to the activities specified in Figure 1-1, of the remainder of this study consists of the chapters specified below.

Chapter 2. Pre-Study

This chapter describes the initial research idea that focuses on mining techniques in the field of information systems. Therefore, we begin by identifying the abundant elements (data, information, knowledge, documents and texts, among others), and the object of the mining process, for which we develop a technique.

Chapter 3. Literature Review about Intention Mining

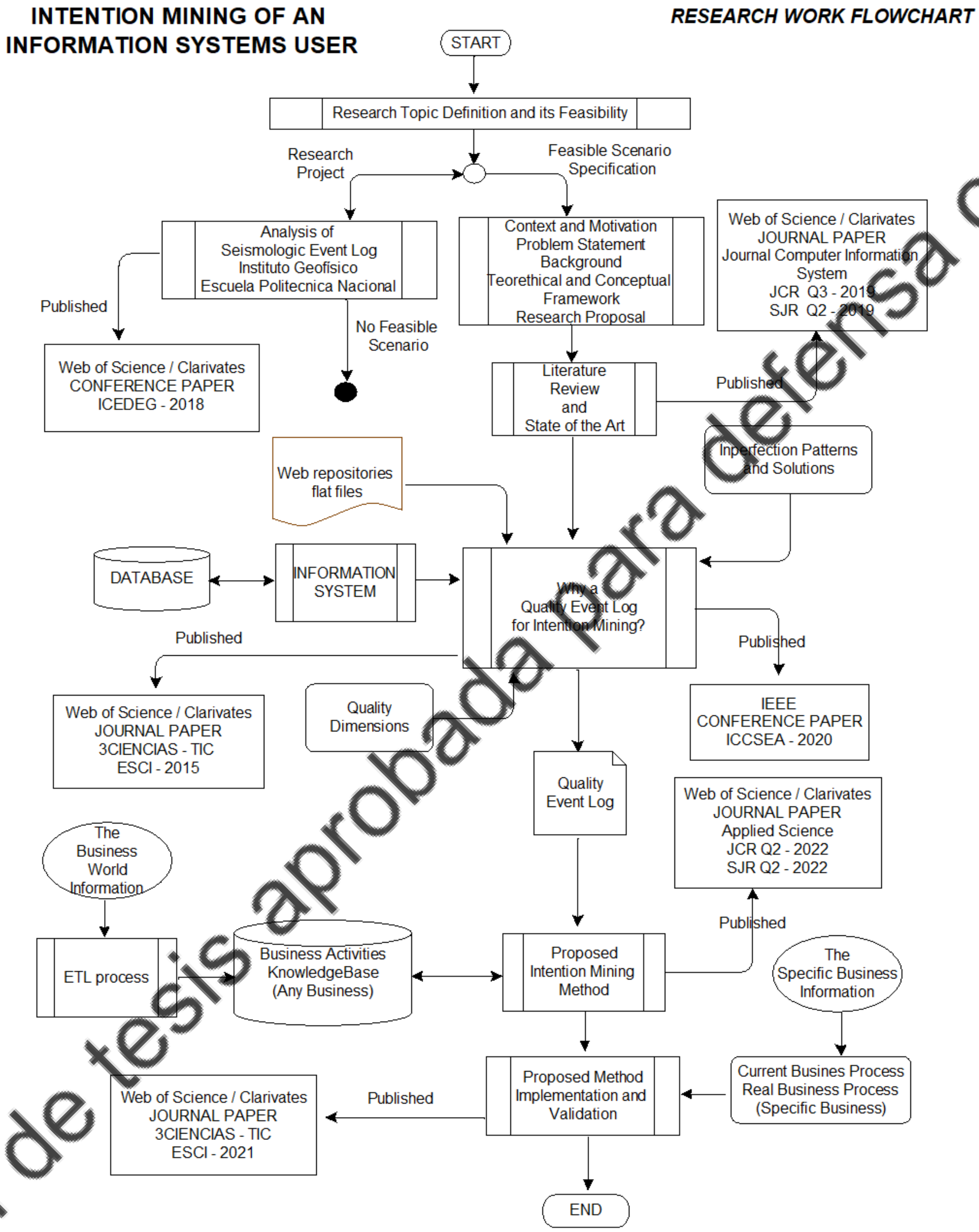
This chapter describes studies that focus on intention mining in an information systems context.

Chapter 4. State of the Art about Intention Mining

This chapter is dedicated to presenting state-of-the-art research [44] on the fundamental topics of mining such as data mining (traditional methodologies), process mining (business application), opinion mining (discussion forums) and text mining (text processing), speech). However, it mainly addresses research on intention mining, the existing methods, tools that have been developed and fields that have been explored, etc.

Chapter 5. Why a Quality Event Log to Intention Mining?

This chapter presents an initial essay on intention mining based on a case study. In the previous chapter, the need for quality information to obtain a record of quality events is addressed. For this reason, this chapter reviews the method for designing a relational database that guarantees the consistency of data to generate high-quality information, proposed by the author of this thesis in [5].



Reserch Topic Definition
1. Feasible Scenario Specification (Introduction to the business scenarios)
2. No Feasible Scenario (Pre-Study abouth the seismological scenario)
3. Literature Review
4. State of the Art
5. Wht a Quality Event Log to Intention Mining?
6. Proposed Mining Method
7. Proposed Method Implementation and Validation

Figure 1-1. Flowchart of Methodology and Research

Chapter 6. Proposed Method to Mine the User Intentions of Business Information System.

According to traditional mining methodologies (data mining), we attempted to conceive a method for conducting intention mining based on general knowledge from the business world (knowledge base) and the information of a specific business (log). Moreover, by using text mining and natural language processing (NLP) techniques, we intended to develop a method that can extract the intentions of users to identify and define improvements in business information systems.

Chapter 7. Proposed Method Implementation and Validation

The method developed in the previous chapter is applied to a sales business to validate its operation.

Chapter 8. Result Discussion

Chapter 9. Summary

CHAPTER 2

2. PRE-STUDY

This chapter is based on the previous publication [32]. In Section 2.1, an introduction to the seismology events of this scenario and their treatment with the mining tools (data, process, text and opinion) is provided. Section 2.2 describes the current situation of IG-EPN seismology area. In Section 2.3, a brief summary of research project development is provided. Section 2.4 presents a discussion of the obtained results, and conclusions are presented in Section 2.5.

2.1. Introduction

The problem that this doctoral thesis attempts to solve was first addressed by a research project in the field of seismology, but over time it was found that the area of intention mining could not feasibly be applied to this field without the use of event log records. This project focuses on creating the necessary information to apply mining in a seismology area of the Geophysical Institute of the National Polytechnic School (called, IG-EPN). Therefore, it is essential to first create a user event log record for the development of the activities through an information system.

2.2. Seismology Area of the IG-EPN

The IG-EPN is the main entity dedicated to determining seismic and volcanic risks in Ecuador, which, together with international teams, is responsible for establishing systems to continuously monitor and record seismic and volcanic activity. Thus, the analog signals (time series) produced by seismic sensors installed on volcanic terrain are processed with the SeisComp3 tool [45] in order to obtain discrete and tabular data in real time that are stored and analyzed using data mining techniques. Using SeisComp3, the data record of seismic events can be recorded in a MySQL database. Expert seismologists generate their reports using Quake ML [46] and tools provided by the Incorporated Research Institutions for Seismology [47].

2.2.1 Acquisition, preprocessing and storage of seismic data

Firstly, the time series recorded by the sensor equipment (installed in the ground by specialized systems that digitize them and generate the discrete data) are processed. The SeisComp3 tool [45] through its functionalities (Scautopick, Scautoloc, Scamp, Scmag, Scevent, etc.) extracts data from seismic events and stores them in a MySQL database. In addition, it generates recorded data in the format of mSeed [48], as shown in Figure 2-1.

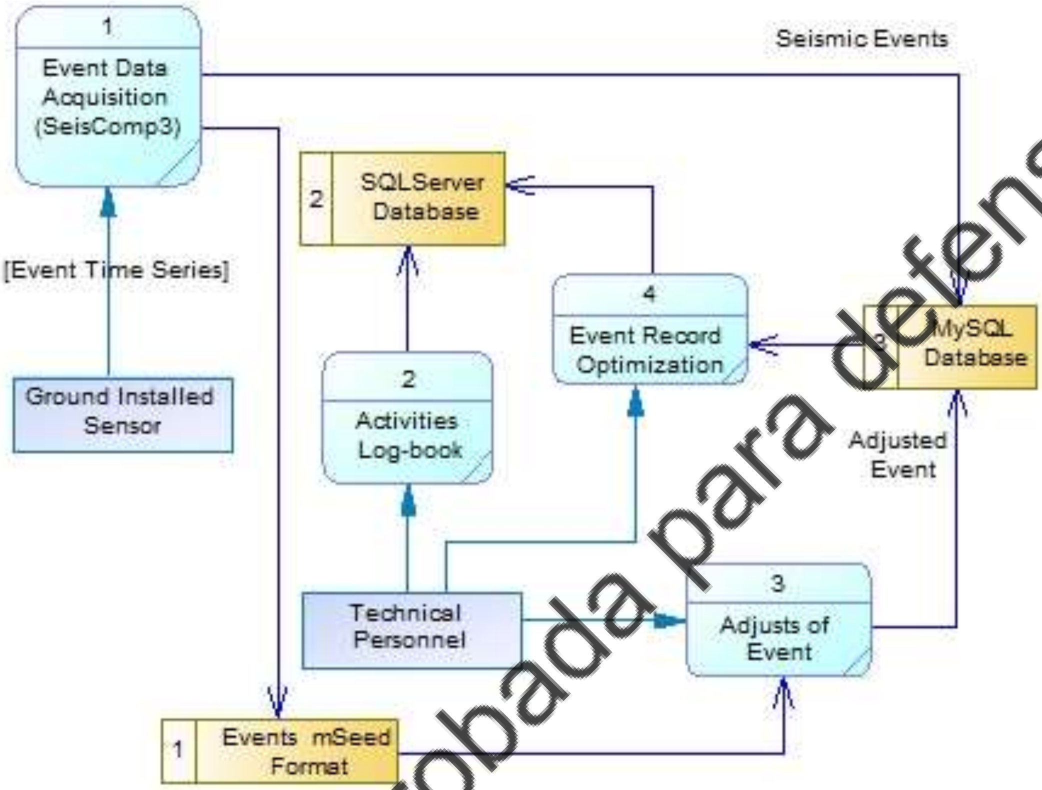


Figure 2-1. IG-EPN Seismic Data Acquisition DFD.

Based on the data pre-processed and stored in the MySQL database through the functionalities and benefits offered by [45], SC3 Forum (Sclob, Scboletin, etc.) and Gempa Dissemination Server, specialized technicians and expert users (in the area of seismology) generated bulletins, reports, and other information to disseminate in the community. We recommend the implementation of the process shown in Figure 2-2.

QuakeML [46], SeedLink, ArcLink, aiUtils, and WebDC3 are the technological resources available in the IG-EPN. Technicians and experts use them to carry out the daily seismology activities. However, for the purpose of this study, the most important resources are the maintenance of the logbook, the recording of adjusted seismic events, and the preparation and issuing of reports of seismic events by seismology experts.

Therefore, from the execution of these activities, we obtained data that feed the log of seismic events. The data used to generate the records of the seismic event logs were obtained by applying process mining to the log book, plus the process log generated by the [45] and the MySQL database storage. Therefore, we also recommend generating a record of seismic events by applying text and opinion mining to the reports issued by seismology experts.

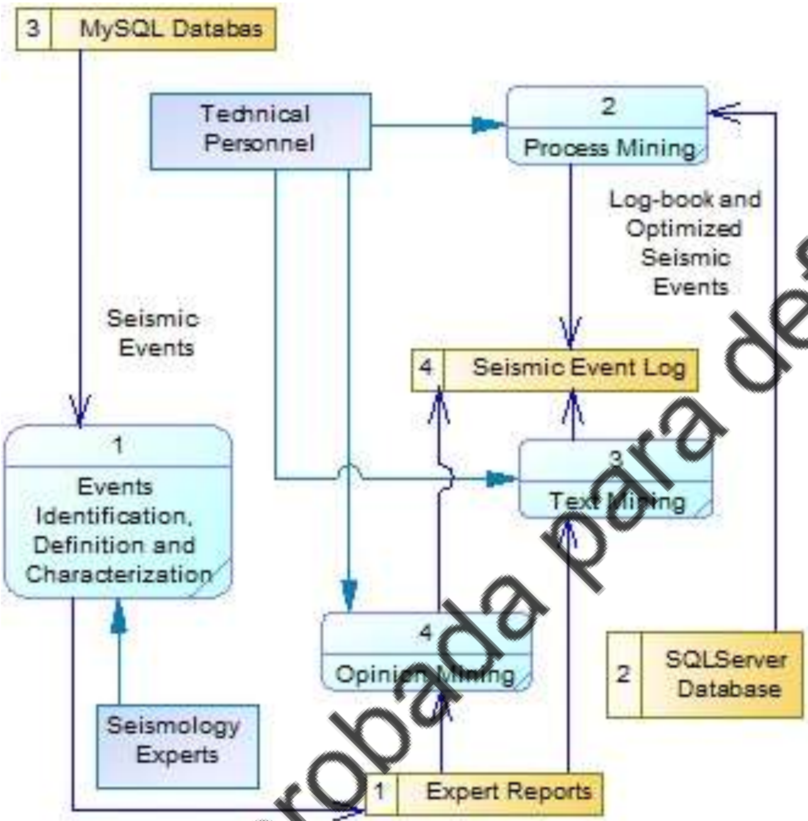


Figure 2-2. Seismic Event Log Generation DFD

2.3. Method

The activities developed in accordance with the planning of the PII-DICC-01-2018 (IG-EPN) project, are a survey of the current situation, definition of the scope and limitations, definition of the problem, analysis of possible solutions, and conclusions and recommendations. Periodic work meetings were held, data were verified, and reports were prepared.

2.3.1 Finding Mining Method

Currently, in IG-EPN, the modeling of seismic events is carried out, for the most part, through data mining tools. We also recommend using process, text and opinion mining tools:

- Data Mining: All the data captured from the environment (scene of the seismic events), are processed using data mining techniques through SeisComp3.

- Process Mining: Process mining techniques allow the extraction of knowledge from process registers that are available in information systems [48]. The use of these techniques can extract patterns in the processes determining the structure of the log, which supports the design of the log for seismic events.

- Opinion mining: For Liu [49], among the other issues that we have addressed, the extraction of opinions is based on features used to characterize seismic events. In this way, the extracted characteristics of the events specified in the reports of seismological experts create the data and metadata of these events, which contribute to the design of their log. In addition, we have considered developing ontology to characterize seismic events in future research.

- Text Mining: We used information extraction techniques [50], [51], [52] from PDF documents (seismic expert reports), previously classified in the present study for the identification of seismic events by seismology experts.

2.3.2 Selection of Methods

For data collection, the SeisComp3 system (Figure 2-1) was used on a daily basis. This software tool is based on data mining techniques. For the development of the activities of technical personnel and seismology experts, the use of process, text and opinion mining tools was recommended (Figure 2-2).

Previous studies have presented several available models: map meta-model formalism, map process model, intentional process model, intentional process meta-model and discover process model. In all of these studies, techniques of process mining and intention mining were developed and applied to the event log, which is generated daily by the information systems. We recommend designing a seismic event log that can identify the differences between the seismic events characterized by the experts in the IG-EPN and real events. The identification of such differences is a topic for future research, which could be developed via an intention mining method.

2.4. Results and Discussion

In this study, the possibility of addressing the topic "intention mining" in the area of seismology is analyzed, where the main protagonists is the seismological event. We focus on its occurrence and the possibility of establishing behavior patterns based on activities, strategies and intentions.

2.4.1 Technical Features

Based on the systemic approach [40], the universe could be thought of as a macrocosm, Earth as a mesocosm, and the human being as a microcosm. Then, universal laws are replicated on Earth and consequently for human beings. In this sense, it would be technically possible to define the intentions of planet Earth (as is done with the human being), but at this time, it is not known that such a possibility exists.

2.4.2 Technological Features

From the identification of the current situation of the seismology area, there are tools, opinions, and texts that can help perform data mining processes [7], [45], there are no tools for carrying out intention mining; that is, there is no possibility of identifying the strategies (intentions) of planet Earth and the occurrence of seismological events.

The applied technique to the prepared report repository (PDF documents issued by seismology experts) enables the extraction of information [51], [52] that identify entities, events and their within semi-structured and unstructured texts. Several software tools could be used, such as text.IE, WordStat, Textalytics, and T-LAB, etc.

Text mining can extract patterns from reports (issued by seismological experts) that facilitate the definition of a report model in terms of its structure and content. Based on this model, we define part of the information that contributes to the design of the new seismic event log.

As shown in Figure 2-1, the technical personnel in the IG-EPN develops daily activities (which are recorded in the logbook) based on the data generated by [45] in mSeed format and with the tabulated data that are saved in the MySQL database. In addition, the technical personnel make adjustments to the seismic event records in mSeed format, and these adjusted records are also stored in the MySQL database. The seismology experts (Figure 2-2) then issue their reports based on MySQL database information. From the current information technology situation that is handled in the IG-EPN, the data structures for version 1 of the seismic event log design are extracted.

The feature extraction [49] of seismic events is a technique of opinion mining applied to the repository of reports prepared and issued by seismology experts. For the execution of opinion mining, we use the R language methodology offered by the University of Virginia [50] that allows, first of all, to build a vector with PDF documents and then extract their characteristics using the products of this methodology (PdfInfo, PdfToText, and PdfTpText). Using this methodology, we can also carry out text mining.

2.4.3 Operative Features

It is possible to establish with some precision when and where seismological events can occur. In contrast to the information system user, their activities (place, date, duration, etc.) could be recorded in an event log.

2.4.4 Operational Features

By applying opinion mining to the reports of the experts, patterns of ideas, thoughts, estimates, polarities and intentions that could contribute to the design of the new seismic event log can be obtained.

2.5. Conclusion

The project developed in the seismology area of the IG-EPN highlights the non-feasibility of this scenario for the development of research on "intention mining". Therefore, I aim to define the approach to this subject in the business world.

CHAPTER 3

3. LITERATURE REVIEW ABOUT INTENTION MINING

This chapter is based on the previous publication [53]. Section 3.1 provides an introduction to mining techniques; Section 3.2 presents the objectives and related literature reviews in an information system context; Section 3.3 shows the proposed method for conducting a literature review; Section 3.4 presents the results obtained by applying the proposed method; Section 3.5 presents the obtained results; Section 3.6 describes implications and future research; and finally, Section 3.7 presents the conclusions.

3.1. Introduction

In the context of information systems engineering (ISE), event logs have become the most important repository of data, information and business knowledge [7]. An event log is the input for the development and application of mining techniques; nowadays, they are undoubtedly the most valuable tools in this field. These tools allow for the alignment of modern technological architectures with business objectives, which are based on the improvement of process models and projection to future cases of business environments, which could be improved through the application of intention mining techniques [19]. In every business, it is possible to demonstrate the presence of data, information and processes that are executed on a daily basis by users at all levels (management, control, operational). As a result of the execution of business processes, the event log [8], [54] is automatically generated or previously configured by technical staff. In this event log, all daily business activities are recorded. The event log, until today, has been used to discover actual processes of business through process mining. We aim to determine the current state of the development of intention mining in an ISE context, so that in future research, we can propose the development of a new intention mining technique that enables the extraction of user strategies from event logs (user intentions) of their daily activities.

In the present study, a method is proposed for conducting a literature review on a new topic, "mining of intentions", regarding the users of the information systems and from the event log, which by default is generated by all information systems.

3.2. Aims and related research

The processes extracted from event logs using process mining are the actual business processes that are executed daily. There is evidence that these actual processes are different from current business processes [7], so a literature review is urgently required to establish state-of-the-art intention mining research. The difference between current processes and actual processes could be caused by users developing their daily activities based on current processes, as well as users utilizing their own strategies (intentions).

This motivated us to develop a research project "Intention Mining Method for identify the differences between Business Current Processes and Actual Processes". This project is part of the doctoral program in computer science at Escuela Politécnica Nacional (Quito, Ecuador). There is a vast number of literature reviews in the field of health sciences, with Cochrane as their principal representative [67]. On the other hand, especially in the field of information technology and an ISE context, several authors have proposed literature review methodologies during the last decade [55], [56], [57], [58], [59]. These studies form the basis of this study. The contributions of every valuable study are summarized in the following.

3.2.1 General method for Cochrane reviews

It is known that the Cochrane protocol is broadly applied in the field of health. It comprises the following activities: Defining the review question and developing criteria for including studies, searching for studies, selecting studies and collecting data, assessing the risk of bias in the included studies, analyzing data and undertaking meta-analyses, addressing reporting biases, presenting results and “summary of findings” tables, interpreting results, and drawing conclusions. We followed every activity of this method in the construction of our own method for a literature review in the field of ISE.

3.2.2 An eight-step guide to conducting a literature review

Okoli and Achabram consider eight steps: Identify the purpose, draft a protocol, train the team, conduct a practical screening, search the literature, extract data, appraise quality, synthesize the studies and write the review. We considered that the search of the literature and appraising quality step helps in our evaluation process.

3.2.3 Literature review: processing

This is a non-systematic review study; nevertheless, we consider that the method of extracting knowledge from the reviewed documents is useful for our purposes [55].

3.2.4 A literature review on strategic information systems planning

This literature review proposes the following activities: Identifying the list of journals; searching for relevant papers; retrieving the first list of papers; excluding papers based on their titles, abstract, and full texts; retrieving the final list of papers; putting appropriate labels on papers; and identifying focus/context/topic and limitations of the research. For our purpose, it is useful to "define the scope and context of the review" [57].

3.2.5 Software engineering—a systematic literature review

This method comprises research questions, searches, inclusion and exclusion criteria, a quality assessment, data collection, data analysis and deviations from protocol: the search results, quality evaluation of the literature review, and quality factors. The Discussion section explores the current leading research topics and their limitations. This study allows us to fine-tune the process of defining research questions [58].

3.2.6 Guidelines for performing systematic literature reviews in software engineering.

This guide applies to the area of software engineering. Yet, for ISE, aspects such as the definition of the structured research questions, search strategy for documents, and criteria for the selection and evaluation of documents [57], [58] are considered to be the contextual base for the development of our literature review method. Based on [58], we propose a method adapted to the ISE context and applied to case studies of interest, i.e., intentional process mining and intention mining [59]. There are proposed solutions that go from data mining to process mining to establish data patterns and improve processes. Given the need to incorporate flexibility into processes [60] and model designs for a process-aware implementation in information systems, a new approach called "intention mining" has emerged [15]. Its objective is to model intentional process representations that can expand process-aware information systems. This also allows users to achieve business objectives by executing processes according to their strategies and intentions without losing track of business rules and current processes (procedure manual).

There are a vast number of papers that use intention mining techniques and tools in areas other than business information systems, such as the following: detecting intentions in questions addressed to users of a specific product or service; the automatic analysis of intentions behind public tweets on the topic of heath [61]; building recommendations based on the remarks and discussions of software developers [62], [63]; the learning and classification of real-time intention expressions contained in microblogs [64]; building a map of intentions of web users to improve the results of search engines [65]; and identifying multilevel intentions in surgical and resuscitation medical treatments (a case study) [66].

Generally speaking, in every business, information systems automatically generate event logs [54] that contain detailed information about business daily processes. Process mining techniques are applied to event logs [7] to extract the actual steps of business processes. These processes are performed based on current processes, plus users' strategies [15], [21]. Consequently, actual processes and current processes are different, but this does not mean that business objects are not accomplished. Nevertheless, there is no evidence for the improvement or evolution of business processes [19]. On the other hand, the emission of runtime recommendations through intention mining techniques have been proposed to gain improvements in those processes [60].

3.3. Proposed Method

Based on Cochrane's [67] proposal for systematic reviews of interventions [58] Systematic Literature Review Methodology, and [59] Guidelines for Performing Systematic Literature Reviews in Software Engineering—we propose a method for developing a literature review in an ISE context [53]. The proposed method consists of the following steps.

3.3.1 Research question

In this step, we formulate the research question in a free format.

3.3.2 Structured research questions

Here, we specify our research questions and ensure that they include the population to which the research is addressed. We determine the procedure, method or protocol [67], which serves as an axis of the literature review, to provide answers to the research question, research topics, and the expected results of the literature review, which support the answers to the research questions in the defined context.

3.3.3 Search process

In this step, we structure search chains formed by a combination of research topics for every expected result. Searches are performed in search engines that belong to each data source. Advanced searches are configured by considering the following elements: research topic keywords and expected results, determined using documents or a full search chain. We also specify research areas, sub-areas and sub-sub-areas (where possible). Additionally, we specify a temporary search window (start and end expressed in months/years).

3.3.4 Document selection

In this step, non-duplicated documents that address at least one research topic are selected. They must be written in English or Spanish and ranked in Journal Citation Reports or SCimago Journal Rank or Proceedings according to the CORE Ranking Portal 2018 [68], which includes Proceedings A*, A, B, C, D, etc.

3.3.5 Assessment of the quality of the selected documents

Here, we established a document ranking system for each of the following evaluation questions (EQ). As a result of this ranking, we obtained a set of relevant documents.

EQ1. Does the document address the research topics that are objectives of this literature review?

1 point: it addresses one topic;

2 points: it addresses two topics;

3 points: it addresses three or more topics.

EQ2. Are the expected results being applied in case studies or developed solutions?

1 point if one expected result is applied;

2 points if two expected results are applied;

3 points if three or more expected results are applied.

EQ3. Does the document specify related studies?

1 point for every related work; maximum of three points.

EQ4. Is their own methodology used in study development?

1 point, if the methodology is unfamiliar;

2 points, if the methodology is of their own.

EQ5. Does the document present an analysis of the expected results?

1 point for each expected result that is analyzed; maximum of three points.

EQ6. Does the document present an application of case studies?

1 point for each case study; maximum of three points.

EQ7. Does the document present an evaluation of the developed study?

1 point if evaluated.

EQ8. Does the document suggest developing future studies?

1 point for each suggested future work; maximum of three points.

3.3.6 Synthesis of relevant documents

For each relevant document, we synthesize: methodologies, techniques, methods, algorithms, procedures, architectures, reference frameworks and standards that have already been developed and used in the documents. The aspects of this step are very relevant according to the research area, research topics and expected results, as defined by the researcher.

3.3.7 Findings extraction

Based on the analysis of the documents, we identify data, information and knowledge about the research topics, expected results and relevant aspects in the literature review.

3.4. Results of Method Application

Once the context, objectives and research scope were defined, we used our proposed method to develop the literature review.

3.4.1 Research question in a free format

The research question (RQ) is defined as follows: What is available in the current literature on intention mining in an ISE context?

3.4.2 Structured research questions

The free-format RQ was formulated in an information technologies context in the ISE domain. In this regard, we defined the following research topics: intention mining and intentional process mining. We also defined the following expected results: methodologies, algorithms, models and tools. Based on the free-format RQ, we devised the following structured research questions:

RQ1. What was developed in the fields of intention mining and intentional process mining between 2010 and 2018?

RQ2. What methodologies have been developed to perform intention mining and intentional process mining?

RQ3. What algorithms have been written to perform intention mining and intentional process mining?

RQ4. What models have been designed to perform intention mining and intentional process mining?

RQ5. What tools have been built to perform intention mining and intentional process mining?

3.4.3 Search process

To define the search process, we specified the temporary window as ranging from January 2010 to December 2021; the research area is set to information technologies, the subarea to ISE, and the research subarea to information systems. Hence, the following search chains (SC) were structured:

SC1: Intention Mining Methodology;

SC2: Intention Mining Algorithm;

SC3: Intention Mining Model;

SC4: Intention Mining Tool;

SC5: Intentional Process Mining Methodology;

SC6: Intentional Process Mining Algorithm;

SC7: Intentional Process Mining Model;

SC8: Intentional Process Mining Tool.

By applying this search process to the extraction sources, as specified in Table 3-1, we retrieved 1965 papers, as shown in Table 3-2.

Table 3-1. Extraction source of documents

Name	URL
ACM	http://dl.acm.org/
IEEE XPLORER	http://ieeexplore.ieee.org/
SCIENCE DIRECT	http://www.sciencedirect.com/
SCOPUS	https://www.scopus.com/
SPRINGER	http://www.springer.com/
TAYLOR & FRANCIS	http://taylorandfrancis.com/
WEB OF SCIENCE	https://login.webofknowledge.com/

Table 3-2. The quantity of extracted documents

SEARCH CHAIN	ACM	IEEE XPLORER	SCIENCE DIRECT	SCOPUS	SPRINGER	TAYLOR & FRANCIS	WEB OF SCIENCE	TOTAL
SC1	7	2	0	4	58	33	39	143
SC2	23	2	2	5	78	19	143	272
SC3	24	2	2	6	95	54	284	467
SC4	14	2	0	4	78	49	62	209
SC5	1	92	0	0	10	18	2	123
SC6	2	186	0	4	15	12	18	237
SC7	2	213	0	4	18	33	34	304
SC8	2	150	0	3	16	34	5	210
TOTAL	75	649	4	30	368	252	587	1965

3.4.4 Selected documents

According to the current proposed method, the selected papers are shown in Table 3-3.

Table 3-3. Selected documents

ID	TITLE	AUTHOR
D1	A Model of Perceptual Task Effort for Bar Charts and its Role in Recognizing Intention	Stephanie Elzer (2006)
D2	A novel approach to process mining: Intentional process models discovery	Khodabandelou Ghazaleh (2014)
D3	A Practical Extension of Web Usage Mining with Intentional Browsing Data Toward Usage	Yu-Hui Tao (2008)
D4	Classification of Switching Intentions Toward Internet Telephony Services: A Quantitative Analysis	Sung Ho Ha (2012)
D5	Contextual recommendations using intention mining on process traces: Doctoral consortium paper	Khodabandelou Ghazaleh (2013)
D6	Customer Revisit Intention to Restaurants: Evidence from Online Reviews	Xiangbin Yan (2013)
D7	Detection of real-time intentions from micro-blogs	Nilanjan Banerjee (2014)
D8	Dynamic Query Intent Mining from a Search Log Stream	Yanan Qian (2013)
D9	Exploring managers' intention to use business intelligence: the role of motivations	Yu-Wei Changa (2014)
D10	Intent Mining in Search Query Logs for Automatic Search Script Generation	Chieh-Jen Wang (2014)
D11	Intentional Process Mining: Discovering and Modeling the Goals Behind Processes Using Supervised Learning	Rebecca Deneckere (2014)
D12	Mining and Ranking Users' Intents Behind Queries	Pengjie Ren (2015)
D13	Mining Coordinated Intent Representation for Entity Search and Recommendation	Huizhong Duan (2015)
D14	Mining Product Intention Rules from Transaction Logs of an Ecommerce Portal	Ravi Chandra Jammalamadaka (2009)
D15	Mining Users' Intents from Logs	Khodabandelou Ghazaleh (2015)
D16	Process mining versus intention mining	Khodabandelou Ghazaleh (2013)
D17	Supervised intentional process models discovery using Hidden Markov models	Khodabandelou Ghazaleh(2013)
D18	Understanding User Intent on the Web Through Interaction Mining	Loredana Caruccio (2014)
D19	Unsupervised Discovery of Intentional Process Models from Event Logs	Khodabandelou Ghazaleh (2014)
D20	User Intention Modeling in Web Applications Using Data Mining	Huan Liu (2002)

D21	Using shared representations to improve coordination and intent inference	Joshua Introne (2006)
D22	Web Usage Mining with Intentional Browsing Data	Yu-Hui Tao (2008)
D23	What shall I do next? Intention mining for flexible process enactment	Elena V. Epure (2014)

3.4.5 Evaluated documents

The high-quality results for each document are presented in Table 3-4. We can conclude that the most relevant documents that contribute to answering our research questions are D2, D5, D11, D15, D16, D17, D19 and D23.

Table 3-4. The quality assessment of documents

ID	EQ1	EQ2	EQ3	EQ4	EQ5	EQ6	EQ7	EQ8	TOTAL
D1	0	0	0	0	0	0	0	0	**0**
D2	1	1	3	2	3	1	1	1	**13**
D3	0	0	0	0	0	0	0	0	**0**
D4	0	0	0	0	0	0	0	0	**0**
D5	1	3	3	2	1	1	1	0	**12**
D6	0	0	0	0	0	0	0	0	**0**
D7	1	0	0	0	0	0	0	0	**1**
D8	0	0	0	0	0	0	0	0	**0**
D9	0	0	0	0	0	0	0	0	**0**
D10	0	0	0	0	0	0	0	0	**0**
D11	2	2	3	2	3	1	1	3	**17**
D12	0	0	0	0	0	0	0	0	**0**
D13	0	0	0	0	0	0	0	0	**0**
D14	0	0	0	0	0	0	0	0	**0**
D15	2	2	2	2	3	1	1	3	**16**
D16	1	3	0	2	2	1	0	0	**9**
D17	2	2	2	2	3	1	1	3	**16**
D18	0	1	0	0	0	0	0	0	**1**
D19	2	2	3	2	3	1	1	1	**15**
D20	0	0	0	0	0	0	0	0	**0**
D21	0	0	0	0	0	0	0	0	**0**
D22	0	0	0	0	0	0	0	0	**0**
D23	2	2	3	1	3	1	1	3	**16**
TOTAL	**14**	**18**	**19**	**15**	**21**	**8**	**7**	**14**	**107**

3.4.6 Synthesis relevant documents

The relevant papers, as well as their methodologies, algorithms and processes, models and meta-models, and software tools are shown in Table 3-5.

3.4.7 Extracted Findings

In Table 3-5, the most important findings regarding data, information, and knowledge (concepts, case studies, and suggested future research) and the most important authors from this field of literature reviews are specified.

Table 3-5. Most used resources in the current reviewed literature

DOCUMENT	D2	D5	D11	D15	D16	D17	D19	D23
METHODS AND TECHNOLOGIES								
Mylyn, Equinox, team/CVS, Junit, OSGI	x			x		x		
Tropos			x					
Machine Learning		x	x	x	x		x	
Process Mining	x	x	x	x	x	x	x	x
Maximum-Likelihood Estimation	x		x	x	x	x		x
Map Miner Method (MMM)	x			x		x	x	
MAP	x	x	x	x	x	x		x
ALGORITHMS AND PROCESS								
Weaver	x							
k-means	x	x		x		x		
Petri-Nets	x	x	x	x	x	x	x	x
Baum Welch	x	x		x	x	x		
Alfa		x		x		x	x	x
Directed Acyclic Graphs	x	x		x	x			x
Hierarchical Clustering		x		x	x			x
Instance Graph		x			x			x
Genetic	x	x		x	x		x	x
Rnet, Inductive Workflow Acquisition	x	x		x	x			x
Ktail	x			x	x	x		
Heuristic Algorithm	x			x		x	x	x
Deterministic Finite State Machine					x			
Support Vector Machine		x	x		x			
KAOS, I*		x	x		x			x
MODELS AND META-MODELS								
Hidden Markov Model	x	x	x	x	x	x	x	x
Naive Bayes Classifier			x					
Dynamic Bayesian Network			x	x			x	
Technology Acceptance Model	x			x				
SOFTWARE TOOLS								
ProM	x					x	x	x

Eclipse Usage Data Collector, Language ToolKit	x			x		x		
Snare		x						
F-Score			x		x			
F-Measure	x			x	x	x		
XES Standard							x	x

3.5. Discussion

We analyzed the relevant aspects and created some criteria from the experience of the literature review performed in this study.

3.5.1 How much did intention mining develop between 2010 and 2018?

We defined the temporary window as between 2010 and 2018; however, we observed that the development of this novel research field, "intention mining", is relatively recent. It started in 2013 with the following publications: [21], [76], [76], [9]. In particular, Ghazaleh Khodabandelou presented the initial exploration of this field in their publication "Contextual Recommendations using Intention Mining on Process Traces" [19]. In addition, it is worth mentioning that "Process Mining Manifesto" [7] is a foundation of the field of intention mining.

3.5.2 Which papers were excluded from the review?

The following papers address the research topics in different ways; however, they were not selected because they do not satisfy selection conditions. For example, these papers cover only one research topic, are not properly indexed, or the conference at which they were presented is not high-ranking:

Process Mining Manifesto [7]. This document was not considered because it is a workshop; nonetheless, it may be considered as the starting point of intention mining as it establishes the process mining principles. Process model discovery checks the process conformance and improves processes. It also presents the following challenges: Finding, merging, and cleaning event data; dealing with complex event records with various features; creating representative benchmarks; dealing with trend change; improving representational bias used for process discovery; balancing quality criteria, such as fit, simplicity, precision and generalization, and inter-organizational mining; providing operational support, combining process mining with other types of analysis; improving usability for experts; and improving understanding for non-experts. These aspects address the importance of event logs in any organization; however, they do not discuss their specialized and specific content depending on the nature, context and area of a business; i.e., every business should have strictly defined structures and content for their event logs', based on the rules, policies and activities of the business.

Intention mining is a solution to parse participant interactions in process-aware information systems [60]. This study was not considered because it is a technical report. However, its contribution has a double intention: Firstly, the authors analyzed the implications for agents, process participants and process administrators and the integrating flexible processes into process-aware information systems through a systematic literature study. Secondly, by using design science, the authors created two artifacts to solve the problem: 1) an innovative process mining technique that discovers the intentional model of the executable process in an unsupervised manner, and 2) a recommendation tool that formulates recommendations as intentions and confidence factors based on partial traces and probabilistic calculus. The artifacts were evaluated in a childcare application case study that supports the flexible process enactment with a data-driven approach. The experiments revealed that the intention mining technique had a precision of 0.69 in discovering the correct intentions. In addition, their intention was to identify the difficulties involved in the development of user activities in an environment that integrates flexible processes into information systems that are aware of the process through process mining techniques. They implement a technique to discover an intentional model of executable processes in an unsupervised way.

Intention-Oriented Process Model Discovery from Incident Management Event Logs [69]. This article was not considered because it is the interpretation of a workshop. However, in its development, it proposes intention-oriented process mining based on the belief that the fundamental nature of processes is mostly intentional (unlike activity-oriented processes) and aims to discover the strategy and intentional process models of event logs recorded during the execution of processes. In this paper, the author presents an application of intention-oriented process mining for the domain of incident management of information technologies. He applied the MMM on a large real-world dataset to discover hidden and unobservable user behavior, strategies and intentions. Although this document is not considered in our literature review, the application developed to discover user strategies and models of intentional processes based on the MMM [9] could be useful in future research on intention mining.

Supervised vs. Unsupervised Learning for Intentional Process Model Discovery [21]. This document was not considered because it is a book chapter, and its contribution is already specified in papers [76], [15]. Learning about human behavior from activity logs requires choosing an adequate machine learning technique regarding the situation at hand. This choice significantly affects the reliability of results. In this paper, Hidden Markov Models (HMM) are used to build intentional process models from the activity log. Since HMMs parameters must be learned, the main contribution of this paper is to compare supervised and unsupervised learning approaches of HMMs.

The MMM is built based on the HMMs and applied to a case study, from which models of intentional processes are generated in supervised and unsupervised manners, the MMM is applied with the purpose of demonstrating the advantages of generating an unsupervised model using the supervised method. In [21], we can see that the intentional process model is generated in terms of user strategies extracted from the events log. This is in accordance with our theoretical foundation to make a literature review.

3.5.3 What are the most used resources in previous works?

The most used resources by previous authors that propose relevant works on the topic of intention mining can be seen in Table 3-5.

3.5.4 Who is leading intention mining research?

As a consequence of the present literature review, we can tell that the most significant authors working on this novel approach called "intention mining" are: Ghazaleh Khodabandelou; Charlotte Hug; Rébecca Deneckère and Camille Salinesi from the Centre de Recherche en Informatique, Université Paris 1 Panthéon – Sorbonne France; and Elena Epure and Sjaak Brinkkemper from the Department of Information and Computing Sciences, Utrecht University, The Netherlands.

3.5.5 Limitations of this study

The developed method based on Cochrane's protocol [67] and the methodologies of Kitchenham [58] and Keele University [59] enables a literature review to be conducted in an ISE context. The literature review was developed from papers published in JCR and SJR journals and the proceedings of the scientific databases specified in Table 3-1. The aim of the literature review is to identify the development (methodologies, algorithms, models and tools) of intention mining and intentional process mining, as well as their applications, authors, and expectations.

3.5.6 Studies developed in the time window (2019-2021)

Additionally, for this time window and relevant previous studies, the following were obtained and analyzed:

[70]: "From our analysis, the challenges of user intent mining fall into three folds. Firstly, user intent could be express explicitly or implicitly. Implicit user intents do not contain the intent keywords, which is more challenging to classify and recognize users' real ideas. Secondly, research of user intent in many domains is lacking. Thirdly, we also observed that user intent is not stable but changing over time. Intentions could interact with each other and have a time decaying phenomenon. Then how to model this dynamic nature of intention is also important to predict user's interests and information needs". According to this author, the present study deals with the explicit intentions of the user that underlies their strategies. I agree with this author; the research of user intent in many domains is lacking, and the research domain of the present study is business information systems.

[14]: "The user intention mining with respect to business perspective is an important and challenging task due to the varying nature of customer-generated text data. The purpose of this review is to present a brief review of studies pertaining to user intention mining with emphasis on discussing different machine learning and deep learning techniques". These authors developed a literature review to extract the intentions of business users using machine learning and deep learning tools applied to commercial transactions performed by customers on social networks. Our study focuses on the intentions of the user of a business information system.

[62]: "In this paper, we manually categorize 5,408 sentences from issue reports of four projects in GitHub. We propose a deep learning based approach to automatically and more accurately classify sentences into different categories of intentions. A case study on four open source projects with 2,076 issue reports shows that our approach achieves an average of the 68.7%". These authors manually classify sentences (possible intentions) and achieve an average performance of 68.7%. In the present study of news articles, sentences are extracted through a proprietary method of sentence mining using NLP libraries (Manning et al, 2020), and a knowledge base is created for businesses in general, which serves to verify and validate user strategies (intentions), and in the best case, the obtained results suggest a performance of 75%.

[71]: “Today vast and diverse event records of applications exist for almost every scientific domain, making their integration and intelligent exploitation challenging. Intention mining is the ability to predict a user’s goals. Knowing the user’s intention can support the decision-making of the network administrators. The main input of all algorithms used to discover intentional process model is the log file (traces activities), which is unstructured dataset and not ready to be feed as-is to machine learning algorithm. Therefore, this paper aims to describe the data preprocessing steps, which transform the unstructured log file to a structured one”. Similar to the study by these authors, in the present study, the main resource is the event log, which is preprocessed and converted into structured data (quality event log). In addition, the user's intentions are identified, allowing us to determine their behavior, and thus facilitating decision making for business managers.

[72]: “The aim of this work is to conduct a literature review about Intents, Intention Mining and Intent Classification. Nowadays, Intention Mining is widely used in the Information Systems Engineering field. This paper mainly focuses and discusses on the literature review algorithms, models and tools used in Intention Mining. We hope that this information will be useful for developing models to retrieve intentions from the traces of activities and developing various intention mining techniques, which will allow identifying the gaps between the prescribed processes and the actual processes of a business”. This author reviews the literature on "mining of intentions", which is an aspect of the literature review developed in this study.

[73]: “Users use the network more and more frequently, and more and more data is published on the network. Therefore, how to find, organize, and use the useful information behind these massive data through selective means, and analyze user intentions is a huge challenge”.

3.6. Implications and future research

Based on techniques, methodologies, methods, models, algorithms, software tools and relevant papers (see Table 3-5), the previous authors developed MMM, which could be considered the first intention mining tool used to generate intentional process models in the information systems field, i.e., the discovery of users’ strategies (intentions). A new intention mining tool, which is currently under development, can be applied in the identification of the gap between current business processes and actual processes. The steps in the creation of this new tool include: i) The use of process mining to extract user strategies from event logs; ii) defining and structuring user intentions based on their strategies; iii) defining levels of abstraction for the executed processes according user intentions; iv) ranking user intentions according to the levels of process abstraction; v) generating intentional process models using MMM and according to the predefined process abstraction levels; vi) validating intentional process models through this methodology, versus the intentional process models generated through MMM.

On the other side, the experience acquired from this literature review led us to determine that the number of high-quality event logs in information systems is generally low. Hence, we recommend improving the data quality in event logs before applying intention mining to yield quality results. Once we obtain high-quality logs, we will be able to predict user behavior in information systems. Consequently, we can issue improvement and optimization recommendations with respect to business information resources [8].

3.7. Conclusions

Based on the literature review, we can answer the research question (RQ) and research sub-questions, as shown below:

- RQ. What is available in the current literature on intention mining in an ISE context? Answer to RQ: In the eight relevant studies performed between 2013 and 2018, we identified that their authors addressed and provided the starting point of "intention mining", and their contributions are the basis for developing future research in this area.
- RQ1. What was developed in the fields of intention mining and intentional process mining between 2010 and 2018?

  Answer to RQ1: The development of intention mining is quite new; there is evidence that it started in 2013. There are previous articles that address this topic but in contexts different from information systems engineering.

- RQ2. What methodologies have been developed to perform intention mining and intentional process mining?

  Answer to RQ2: Each relevant study has implemented its own methodology. Among the most important are: map miner method, supervised map miner method and unsupervised learning for map miner method. Based on these methodologies, in future research, we intend to build a standard methodology for intention mining in an information systems engineering context.

- RQ3. What algorithms have been written to perform intention mining and intentional process mining?

  Answer to RQ3: According to our literature review, several algorithms have been developed: deep miner algorithm, map miner algorithm, intent-miner, intent-recommender, and strategies miner algorithm. Yet, other tools, such as KAOS, I* and MAP, can also be used as intention mining techniques.

- RQ4. What models have been designed to perform intention mining and intentional process mining?

  Answer to RQ4: Process, strategy and intention modeling will be the main focus of our future research on "intention mining". According to the literature review, a few models that are used in this field are map meta-model formalism, map process model, intentional process model, intentional process meta-model, current process model, and discovered actual process model.

- RQ5. What tools have been built to perform intention mining and intentional process mining?

  Answer to RQ5; currently, in the intention mining of information systems engineering, the following two free software tools that perform intention mining are available: Plug-in ProM ToolKit, and Eclipse Usage Data Collector Language ToolKit. Nevertheless, there are more software tools for intention mining from other data sources and in other areas, such as health and social sciences.

The results of this literature review of a new topic called intention mining triggered the development of new products that can help businesses in agile migrations toward innovative platforms and dynamic solutions according to current and future business goals, such as process flexibility, enterprise resilience, risk management, and a successful decision making.

The most important author (Khodabandelou) in this novel approach proposes the following ideas for future studies: improving guidance through the processes that provide better recommendations for run time, facilitating process modeling, identifying the gap between current business requirements and actual information systems users' goals, helping CIOs in a pro-active way, and monitoring the intentions of users.

Based on this literature review, we are planning to develop an intention mining method to identify the gap between current processes and actual processes of business;

for this, we are designing a methodology that includes new algorithms to enable the automatic generation of intention process models.

CHAPTER 4

4. STATE OF THE ART ABOUT INTENTION MINING

Section 4.1 presents a brief summary of process mining techniques and their application to the business gives a general vision about business; Section 4.2 provides knowledge of the current situation of the intention mining scenario in Section 4.1; Section 4.3 presents the background and development of user intention mining in a business information system context; Section 4.4 synthetizes the findings obtained in the literature review.

4.1. Process Mining

In the survey, verification and validation of business processes, it is common to find that the traces established for the processes do correlate with the activities carried out by the users. However, the aforementioned traces have been elaborated, are the final product of a consultancy study, and have been periodically and permanently updated. The rapid evolution of business—given the constant growth in user demands and the need for real-time responses to information requirements without losing sight of control through the verification and validation of compliance with business rules—is the guideline for the approach of designing new process models. These models should be based on the historical behavior of the business (activity logs), current traces of the processes (implemented in a workflow), patterns obtained from a mining of data, and the profiles of the users and elaborated using the technical and technological tools available. For this, there are process mining tools, which can be applied to a log of business events. Aside from discovering the real business processes that are one of the inputs of intention mining for the objective of this study, there is also the compliance protocol, which establishes the differences in data, forms, procedures, rules, and policies between the current and real processes [7].

4.2. Intention Mining

The development of intention mining has been in the area of social networks is extensive. In the context of information systems, from the year 2013, the main authors (specified in the previous section), taking the event log and the process mining techniques of the author Van der Aalst as main inputs, promulgated the concepts and developed the techniques to mine the intentions of users of the information systems in order to issue recommendations for improving the development of software

4.2.1 Background

Nowadays, the findings, applications and recommendations of the previous authors have become the background for the development of the proposed method to intention mining, as well as the background for the new developments of the subject in question, the business field, its information systems, and its users.

4.2.2 Information Systems Context

In the specific context of our study, the existing applications and study cases identified in the literature review are counted as a baseline.

- Eclipse Usage Data Collector [21]: The purpose of this case study is to obtain a process model map that helps understand the behavior of software developers in terms of their intentions and strategies. The Baum Welch Algorithm (BWA) is used to mine strategies and complete intentions based on these strategies using MMM. The results of this study are a series of recommendations about the strategies that developers must use to complete their intentions. Nonetheless, we believe that using HMMs to extract the strategies would be more effective, since BWA is only applicable to deterministic processes, and developers' behavior is clearly a random process, despite the existence of current processes for the development of their activities.
- Entity/Relationship Diagram [74], [75]: In this case study, the authors provided to a set of students a guide map for the elaboration of an E/R model. The purpose was to establish map sections that are the most used by the students in the elaboration of their model. The method was supervised; hence, the final model contained subjective elements and the database design precision was difficult to measure. That is why we believe that this case study does not significantly contribute to the "intention mining" research field.
- FlexPAISSeer (Flex Process Aware Information System) [9]: This tool can incorporate flexibility in the execution of processes based on knowledge management through the algorithms, "Intent miner" and "Intent Recommender", and with the support of mining technique processes. In their case study, process mining and intention mining techniques are used to incorporate flexibility in a specific process (childcare) through a software tool implemented for this purpose. Identifying and measuring the level of flexibility in these processes through intention mining techniques is quite complex.

Therefore, we considered that it would be more convenient to use other tools, e.g., ontologies and knowledge bases that can analyze and quantify human reasoning in the execution of processes.

- ProM Toolkit Plug-in [21], [74]: Users' strategies can be hidden processes. Process mining tools can represent the processes in terms of activities. These tools have been plugged into the ProM toolkit to offer an intentional vision on processes. This enables the extraction of process knowledge based on process execution logs in an intentional mode. This is the best free tool to process intention mining from process mining techniques; therefore, we will use and recommend this tool for validating the future research.

4.2.3 Intention mining of information systems user

Tables 3-5 (literature review) show the most used resources (methods, algorithms and tools) for the development of previous research in the context and objective of this study. Below is a summary of the most relevant resources, which serve as a reference for our development and that, in some cases, can be used as available tools.

- Intention mining techniques and tools: The intention mining techniques presented in this literature review are based on process mining techniques, HMMs, MMM and other tools described in this study. The most important technique (MMM) is used in the case study of Eclipse Usage Data Collector [21]. In future research, and for this literature review, we intend to develop an intention mining technique to identify the gap between current processes and actual processes.
- Process mining techniques and tools: Generally, they are used as a bridge between data mining and machine learning, to discover process models oriented to user activities, to identify user and HMM strategies, and to determine intentional process models. In addition, process mining techniques could also be used to validate intentional process models [21].
- Machine Learning: This algorithm is used to classify sequences of activities based on similar characteristics. However, a deep machine learning algorithm could also be used to extract the underlying intentions in the traces of the processes. [21].
- Map Miner Method: This is the best product developed so far in this new field of intention mining [21]. It enables intentional process models to be constructed, which could be the basis for future research in this field. Moreover, MMM can define and specify user strategies and activities at different levels of abstraction.

- MAP: This is a modeling language used to build intentional process meta-models that help formalize process models [76]. Nevertheless, we could use UML modeling languages to carry out a similar process, since it is a widely known process in software engineering modeling.
- Petri-Nets: This tool is used to represent process models, particularly, cyclic process models [76]. To achieve our aims, this resource could be used to represent user strategy models.
- Baum Welch: Authors of relevant documents use this algorithm to uncover unknown parameters in the application of HMMs [76]. In our case, we can use it to determine the parameters for mapping user strategies and intentions.
- KAOS: This tool can model requirements as instances of a conceptual meta-model that supports user intentions, but it is not used due to its rigid definition of tasks [76]. Nevertheless, this tool would help to define levels of abstraction in relation to hierarchical divisions of users' activities and strategies.
- Hidden Markov Models: It is used in every relevant document of our literature review to discover intentions from the process traces and elements such as processes, strategies and models, which are related to intention mining and present in the events log. We would recommend using this tool in every future study.
- ProM: This is a free software tool, where the previous authors used their implemented application to create intentional process mining in the development of the case study Eclipse Usage Data Collector. We intend to do the same with the final results of our study [21].

4.3. Scenario Characterization

In the pre-study (chapter 2), a feasible scenario was defined to address the research on the development of the "intention mining" for the users of information systems in a business context. The most relevant characteristics that were defined for the scenario object—the literature review—are described below.

Business Characterization: To identify the current business situation, we proceeded with surveying the processes and determining the activities that are carried out on a daily basis. Based on the information obtained, the feasibility (technical, technological, operative, and operational) of implementing a user intention mining process in the information system is defined.

Business Policies and Rules: For a global approach and in a general business context, a business scenario is delimited by observing the pertinent policies of the business [4]. On the other hand, business rules are defined in a specific way [77], which govern and support its balance according to the level of granularity of the activities and the degree of cohesion in its functions.

Business Process Models: The information system is designed to execute the processes in accordance with the current process models, which are established to achieve business objectives, strictly aligned with business rules. However, users develop their activities based on the aforementioned current process models and their own strategies. This resulted in other (real) process models, proven via the application of process mining in event logs [7], which is automatically generated in the execution of activities through the information system, which we will call generic event log (GEL).

Generalizations: Based on the previously defined concepts, the term "Business" and its application environment, in the intention mining context of this study, can be applied to any business under the following conditions: the business can be defined and characterized (current process) based on the rules and policies of its enterprise; the business activities are developed through an information system; the information system applications can generate the configuration of the event log; and the information system has flexibility, so that the user can use their own strategies in the development of business activities.

4.4. Data, information and knowledge

From the findings obtained in the literature review, the concepts, case studies, suggestions for future research, contributed by the most important authors in this field, are defined.

Concepts

- Process mining techniques are included in data mining and machine learning techniques [19].
- Modeling users' behaviors in terms of activities and ignoring the underlying human cognitive operators, such as intentions and strategies, are the goals of process mining techniques [15].
- User strategies can be hidden processes [21].

- Process mining tools can represent processes in terms of activities. These tools are plugged into the ProM ToolKit to offer an intentional vision of processes. This allows the extraction of knowledge about a process based on process execution logs in an intentional mode [21], [74]. The main contribution of this mining method is to produce intentional process models, i.e., conceptual models of the intentions behind processes.
- In an ISE context, intentional process mining can be useful at different levels of the process model lifecycle: (i) at a requirements level to elicit actual users' goals rather than inferred requirements, (ii) at a project management level to verify whether a current objective and the actual process model correlate, or (iii) at an application level to supervise user activities and provide more useful recommendations during runtime [15], [74].
- Many studies in the field of intentional process modeling have demonstrated that the fundamental nature of processes is mostly intentional; therefore, these processes should be modeled from an intentional point of view. In this regard, modern intentional process models have emerged to offer a flexible structure to model processes [15].
- The main goal of intention mining is to extract sequences of users' activities from event logs to evaluate and predict users' intentions with respect to those activities [76]. Process models only focus on activities, and intentional process models focus on the intentions underlying the activities rather than the activities themselves [76].
- A set of strategies allows users to achieve their intentions and a strategy can be used to achieve several intentions [21].
- The relation between intentions, strategies and activities represents the top-down structure of reasoning and acting in the cognitive processes of the human brain [15].
- A sub-intention is associated with a parent intention, and one intention is fulfilled if at least one of its children sub-intentions is fulfilled [15].
- Several support activities and guidance solutions based on process mining have been proposed, but they lack suitable semantics for human reasoning and decision making. They mainly rely on low-level activities [9].
- Process mining aims to discover, verify the conformance of, and enhance activity-oriented process models from the event log. Intention mining has the same objectives as process mining, but it specifically addresses intentional process models [76], i.e., processes focus on the reasoning behind the activities.

Case Studies

- The map miner method (MMM) is applied to a real-world dataset, which is an event log of Eclipse UDC (Usage Data Collector) developed by Khodabandelou [21] and [78]. The resulting map process model provides a valuable understanding of the processes followed by their developers, as well as feedback on the effectiveness and the demonstrated scalability of MMM.
- A map specifying the intentions and strategies of entity–relationship modeling was given to the students as a guide [74], [76]. In order to obtain traces, they developed a web-based tool that records which sections of the map were followed by the students, while creating an entity–relationship diagram.
- To demonstrate the validity of the FlexPAISSeer approach [9], the previous authors conducted a revealing single case study considering its suitability (the support of flexible processes through its software product). A childcare system was developed by 42windmills and used by several children daycare centers in the Netherlands.
- The ProM framework is a pluggable framework that supports various plugins for different process mining techniques, such as α-algorithm and its extensions [76].

Suggested Future Works

- Intentional process mining might help improve guidance, provide better recommendations, facilitate process modeling and process model quality assessment, identifying the gap between business current requirements and goals and helping CEOs assess and monitor strategic goal implementation [74].
- To establish a base to manually infer the names of strategies and intentions, it can be fully automated by building sophisticated ontologies from the uncovered topics. These ontologies should consider the context in which the processes are enacted as well as current situations [15].
- A ProM plugin for IntentMiner and IntentRecommender must be developed [9]. Official XES extensions must also be proposed to integrate the concepts of process context and entities.

Authors

Ghazaleh Khodabandelou, Elena Epure, Rebecca Deneckere, Charlotte Hug, Camille Salinesi and Sjaak Brinkkemper.

CHAPTER 5

5. WHY A QUALITY EVENT LOG TO INTENTION MINING?

This chapter is based on a previous publication [5], [20]. Section 5.1 provides brief instructions for the event logs in businesses; Section 5.2 presents a method to generate a quality event log (QEL) from a generic event log (GEL); Section 5.3 shows the obtained results for applying this method; Section 5.4 analyzes the obtained results; and Section 5.5 presents the conclusion.

5.1. Introduction

Following the techniques, methods and methodologies of traditional mining (resources to be mined, mining objectives, resource pre-processing, mining tools, obtaining resource patterns and/or predictions [79], and the analysis of the patterns), the resource to be mined in our case was the event log. Our mining goal was to infer user intentions, and for this, it was essential to have a refined event log (quality event log) [80].

The discovery of processes through process mining techniques, from the event log [7], is the starting point for considering the development of an intention mining technique [76] based on process mining techniques. The aim of this study is to achieve a structure for the quality event log (QEL) [81] from the generic event log (GEL). In this way, the QEL can determine user strategies [21] and infer their intentions [32].

5.2. Method

Based on the methods proposed by [80] and [82], the steps of the method to debug the GEL and generate the QEL are presented below.

5.2.1 Obtaining the GEL

We began with the assumption that companies currently carry out their activities through an information system, which processes data stored in databases and, as a product of this processing, the transaction log is generated. To obtain a GEL as a feasible resource to be mined, a good database design must be sought to guarantee consistency in the data, and as a precursor to the event log [5], [83], [84].

The obtained GEL contains records of all the transactions executed in the business: (sales business as a case study [85]; records of current business activities;

records of erroneous transactions; records of errors; records of exceptions; records of user strategies, etc. From the “Sales_log” file, taken from kaggel and contained in GitHub repository [34], Table 5-1 presents an extract of this log that represents the GEL.

Table 5-1. Generic Event Log (GEL)

Trace ID	Event ID	Event Date	Activity	User ID
1	101	2019-11-01 06:18:48	Customer service	274969076
1	102	2019-11-01 06:50:46	Generate customer order	274969076
1	***103***	***2019-11-01 06:52:33***	***Cancel insufficient-stock order***	***274969076***
2	104	*2019-11-01 08:20:16*	Customer service	384989212
3	105	2019-11-01 03:15:33	Customer service	295643776
3	106	2019-11-01 03:12:38	Generate customer order	295643776
3	107	2019-11-01 03:14:51	Local stock control	295643776
3	108	2019-11-01 03:15:36	Dispatch customer order	295643776
3	109	2019-11-01 03:14:19	Register sale	295643776
3	110	2019-11-01 03:14:47	Sales record	295643776
4	111	2019-11-01 14:01:22	Customer service	396477034
4	112	2019-11-01 14:01:33	Generate customer order	396477034
4	113	2019-11-01 14:02:47	Local stock control	396477034
4	***114***	***2019-11-01 23:59:18***	***Register sale***	***396477034***
4	***116***	***2019-11-02 00:00:26***	***Delivery customer order***	***396477034***
4	115	2019-11-02 04:01:27	Home sale delivery	396477034
5	117	2019-11-01 05:06:31	Customer service	296465302
5	118	2019-11-01 04:13:46	Generate customer order	296465302
5	119	2019-11-01 04:13:03	Local stock control	296465302
5	120	2019-11-01 05:07:04	Delivery customer order	296465302
5	123	2019-11-01 05:09:30	Billing custom sale	296465302
6	124	2019-11-01 03:56:15	Customer service	384989212
6	125	2019-11-01 03:56:44	Generate customer order	384989212
6	126	2019-11-01 03:57:08	Local stock control	384989212
6	127	2019-11-01 03:58:38	Complete customer order	384989212
6	128	2019-11-01 04:37:08	Cancel customer order	384989212
7	129	2019-11-01 03:22:57	Customer service	512739566
7	130	2019-11-01 03:20:01	Generate customer order	512739566
7	131	2019-11-01 03:20:20	Local stock control	512739566
7	132	2019-11-01 03:29:04	Complete customer order	512739566
7	133	2019-11-01 03:21:04	Remote stock control	512739566
7	134	2019-11-01 03:22:13	Emit customer quotation	512739566
8	135	2019-11-01 07:15:32	Customer service	512462424
8	136	2019-11-01 07:18:35	Generate customer order	512462424
8	137	2019-11-01 07:21:57	Local stock control	512462424
8	138	2019-11-01 09:54:36	Complete customer order	512462424
8	139	2019-11-01 11:54:34	Remote stock control	512462424

8	140	2019-11-01 13:54:23	Emit customer quotation	512462424
8	141	2019-11-01 13:54:50	Cancel customer order	512462424
9	142	2019-11-01 14:22:30	Customer service	512912806
9	143	2019-11-01 14:30:21	Generate customer order	512912806
9	144	2019-11-01 14:25:15	Local stock control	512912806
9	145	2019-11-01 14:27:16	Complete customer order	512912806
9	146	2019-11-01 14:27:29	Remote stock control	512912806
9	147	2019-11-01 14:29:24	Emit customer quotation	512912806
9	148	2019-11-01 14:29:09	Delivery customer order	512912806
9	149	2019-11-01 14:28:30	Register sale	512912806
9	150	2019-11-01 14:28:47	Sales record	512912806
9	151	2019-11-01 14:29:17	Home sale delivery	512912806
10	152	2019-11-01 18:20:18	Customer service	296465302
10	***153***	***2019-11-01 18:40:19***	***Customer empathy***	***296465302***
10	154	2019-11-01 19:29:20	Generate customer order	296465302
11	155	2019-11-01 20:50:21	Customer service	295643776
11	***156***	***2019-11-01 20:32:22***	***E-commerce omni-channel***	***295643776***
11	157	2019-11-01 21:50:23	Complete customer order	295643776
12	158	2019-11-01 21:09:24	Customer service	396477034
12	***159***	***2019-11-01 21:12:25***	***Customer e-commerce***	***396477034***
12	160	2019-11-01 21:28:30	Deliver customer order	396477034

5.2.2 Current Business Activities Definition

For a sales business (as a case study), from the activities contained in GEL, and based on the files "Sales_policies", "Sales_principles", and "Sales_rules" available in the GitHub repository [34], the activities of the sales business specified below are identified, corroborated and defined.

- Customer service:
    - Create client;
    - Activate client;
    - Update client.
- Generate customer order:
    - Complete customer order;
    - Customer order quote.
- Emit customer quotation:
    - Local stock control;
    - Remote stock control;
    - Register items.
- Complete customer order:

- Local stock control;
- Remote stock control;
- Register items.

- Local stock control:
  - Verify stock;
  - Emit local minimum stock alert;
  - Compromise stock.
- Remote stock control:
  - Verify stock;
  - Emit remote minimum stock alert;
  - Compromise stock.
- Delivery customer order:
  - Dispatched order confirmed;
  - Local partial order dispatch;
  - Emit remote delivery order;
  - Send home sale.
- Cancel customer order:
  - Release committed stock.
- Register sale:
  - Verify confirmed order;
  - Register discounts;
  - Verify taxes;
  - Register payment method;
  - Billing.
- Home sale delivery:
  - Verify customer address;
  - Home delivery tracking;
  - Confirm home delivery.
- Sales record:
  - By product (discount for: liquidation, launch);
  - Per customer (discount for: frequent customer);
  - By warehouse;
  - By period / season;
  - By payment method (current discount);
  - By promotion;
  - By seller;

- By delivery method;
- By volume sales (discount of sale amount).

Based on these activities, the process mining techniques and the ProM tools (alpha algorithm), as well as the sales business current process model, are presented in Figure 5-1.

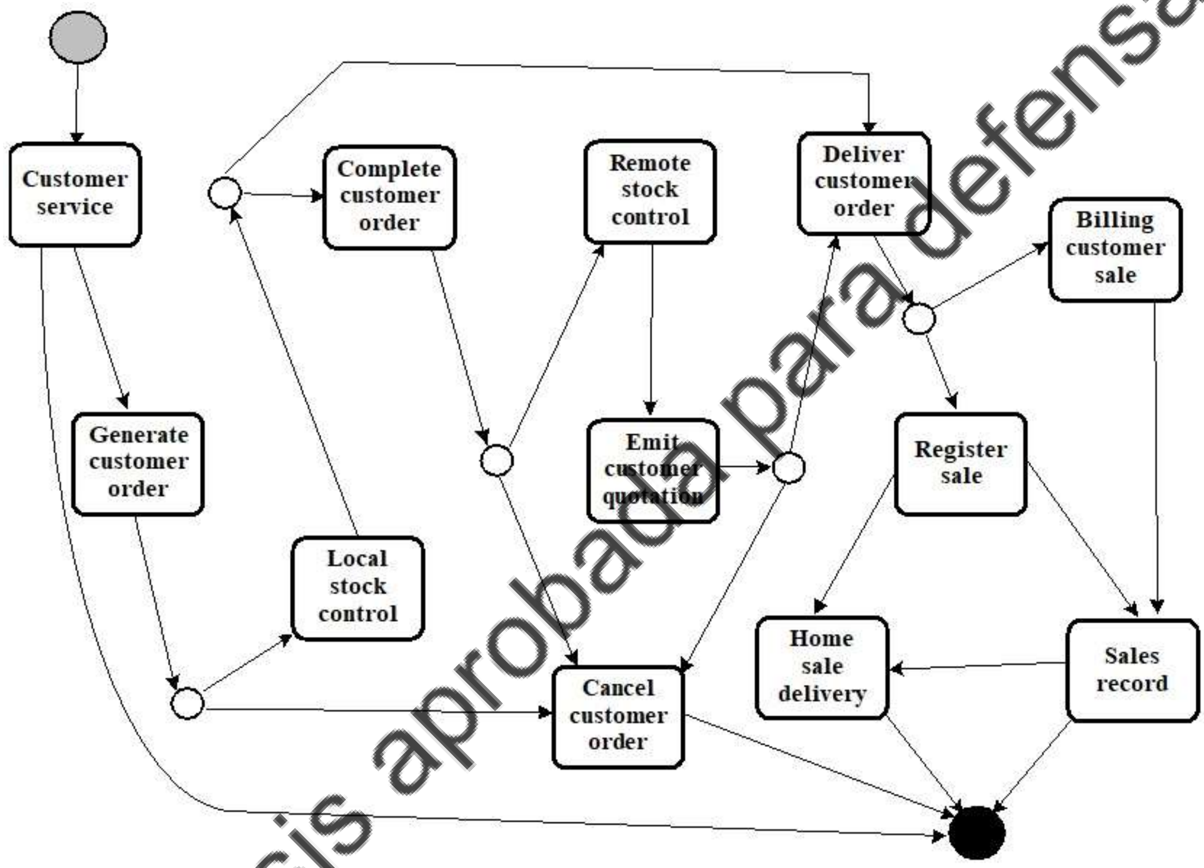


Figure 5-1. Current process model of a sales business.

5.2.3 Quality Dimensions Definition

Taking the ISO Standard 25012 as a reference, the GEL of an information system, whose objective is to provide a data source to verify compliance with the current processes and identify the real processes, which are executed daily in the business, must comply with the quality dimensions (integrity, unambiguity, meaningfulness, correction, syntactic precision, semantic precision). These dimensions allow QEL to be obtained, business rules to be satisfied [86], and the requirements of intention mining to be met. Table 5-2 shows a description of these quality dimensions and their implications in our study.

Table 5-2. Quality Dimensions

No.	Quality Dimension	Description	Implication
1	Integrity [81]	Its structure and content must ensure the existence of a “valid” trace [34] for each and every instance of the real processes, which are executed in the information system.	Inherent in log
2	Unambiguity [81]	Regarding the existence of repeated traces, for the same instance of a real process.	Inherent in business environment
3	Meaningfulness [87]	Any trace in the GEL, which corresponds to a real process instance, is validated through the process model.	Inherent in business environment
4	Correction [81]	Regards the execution time of each and every instance of the real processes; one and only one “valid” trace corresponds to them.	Inherent in business environment
5	Syntactic Precision [87]	Its structure, i.e., the fields of the GEL (sufficient and necessary), correspond to the business domain.	Inherent in log
6	Semantic Precision [87]	Regards the values contained and stored in the GEL and their correspondence with the real values according to the breadth, depth and scope defined for the business domain.	Inherent in log

5.2.4 Imperfection Patterns Identification and their Solutions

To identify the imperfection patterns [35] in the GEL records in accordance with the quality dimensions (Table 5-2), the following imperfection issues were taken as references: data loss (the records are incomplete; empty fields exist); inconsistent data (data are outside the range and do not correspond to the business domain); irrelevant data (data do not contribute to the problem solution); erroneous data (values are unknown or do not correspond to the format, especially in the date values); and inaccurate data (values do not correspond to reality, due to rounding, approximation, truncation or abbreviation). Based on these characteristics, and with the objective of achieving a QEL (objective of the present work), the following imperfection patterns for the GEL were established.

- Chronological disorder of the events:
  This occurs when the process instance starts its execution a little before midnight and ends its execution after midnight [82]. This can cause a certain event to be recorded, with the date of occurrence being the next day (after midnight), but it is saved with the date of the previous day (before midnight). Consequently, the logical order of the events does not correlate with the logical sequence of the process.

  Solution: Adjust the sequence according to the process logic in the process model.

- Differences in the date format between the event log and the tools that process it:
  Normally, the date data contained in the GEL have a pre-established format for the information systems that are executed in the business, but this format date is different from the date format, which is handled in the software tools used to process the GEL.

  Solution: Configure the tools according to the log date format.

- Non-atomic events:
  This occurs when the level of granularity of an activity (contains sub-activities) that corresponds to an event does not correlate with the level of granularity established for the information system according to the business domain.

  Solution: Establish the highest level of granularity in the process model.

- Events that do not correspond to the business domain:
  The activities associated with the execution of the process give rise to the events that are recorded in the GEL, and these activities are not in accordance with the current process model.

  Solution: Verify their correspondence with the user strategies; otherwise, eliminate the event.

- Events that could correspond to user strategies:
  Some events that do not correspond to the business domain activities are validated using user strategies, which will be defined later, in the differences between the current and real process models.

  Solution: Delete the event if it does not correspond to a user strategy.

5.2.5 QEL Generation

Finally, we obtained a quality event log for the sales business, as shown in Table 5-3.

Table 5-3. Quality Event Log (QEL)

Trace ID	Timestamp	Activity
1	2019-11-01 06:18:48	Customer service
1	2019-11-01 06:50:46	Generate customer order
1	*2019-11-01 06:52:33*	Cancel insufficient-stock order
2	*2019-11-01 08:20:16*	Customer service
3	2019-11-01 03:15:33	Customer service
3	2019-11-01 03:12:38	Generate customer order
3	2019-11-01 03:14:51	Local stock control
3	2019-11-01 03:15:36	Deliver customer order
3	2019-11-01 03:14:19	Register sale
3	2019-11-01 03:14:47	Sales record
4	2019-11-01 14:01:22	Customer service
4	2019-11-01 14:01:33	Generate customer order
4	2019-11-01 14:02:47	Local stock control
4	*2019-11-01 23:59:18*	Deliver customer order
4	*2019-11-02 00:00:26*	Register sale
4	2019-11-02 04:01:27	Home sale delivery
5	2019-11-01 05:06:31	Customer service
5	2019-11-01 04:13:46	Generate customer order
5	2019-11-01 04:13:03	Local stock control
5	2019-11-01 05:07:04	Deliver customer order
5	2019-11-01 05:09:30	Billing customer sale
7	2019-11-01 03:22:57	Customer service
7	2019-11-01 03:20:01	Generate customer order
7	2019-11-01 03:20:20	Local stock control
7	2019-11-01 03:29:04	Complete customer order
7	2019-11-01 03:21:04	Remote stock control
7	2019-11-01 03:22:13	Emit customer quotation
8	2019-11-01 07:15:32	Customer service
8	2019-11-01 07:18:35	Generate customer order
8	2019-11-01 07:21:57	Local stock control
8	2019-11-01 09:54:36	Complete customer order
8	2019-11-01 11:54:34	Remote stock control
8	2019-11-01 13:54:23	Emit customer quotation
8	2019-11-01 13:54:50	Cancel customer order
9	2019-11-01 14:22:30	Customer service
9	2019-11-01 14:30:21	Generate customer order

9	2019-11-01 14:25:15	Local stock control
9	2019-11-01 14:27:16	Complete customer order
9	2019-11-01 14:27:29	Remote stock control
9	2019-11-01 14:29:24	Emit customer quotation
9	2019-11-01 14:29:09	Deliver customer order
9	2019-11-01 14:28:30	Register sale
9	2019-11-01 14:28:47	Sales record
9	2019-11-01 14:29:17	Home sale delivery
10	2019-11-01 18:20:18	Customer service
10	*2019-11-01 18:40:19*	Customer empathy
10	2019-11-01 19:29:20	Generate customer order
11	2019-11-01 20:50:21	Customer service
11	*2019-11-01 20:32:22*	E-commerce omni-channel
11	2019-11-01 21:50:23	Complete customer order
12	2019-11-01 21:09:24	Customer service
12	*2019-11-01 21:12:25*	Customer e-commerce
12	2019-11-01 21:28:30	Deliver customer order

5.3. Results

The QEL shown in Table 5-3 is the result obtained from applying this method; it is the quality log that will be used as an input in the intention mining process. Figure 5-2 presents the sales business real process model, built based on QEL.

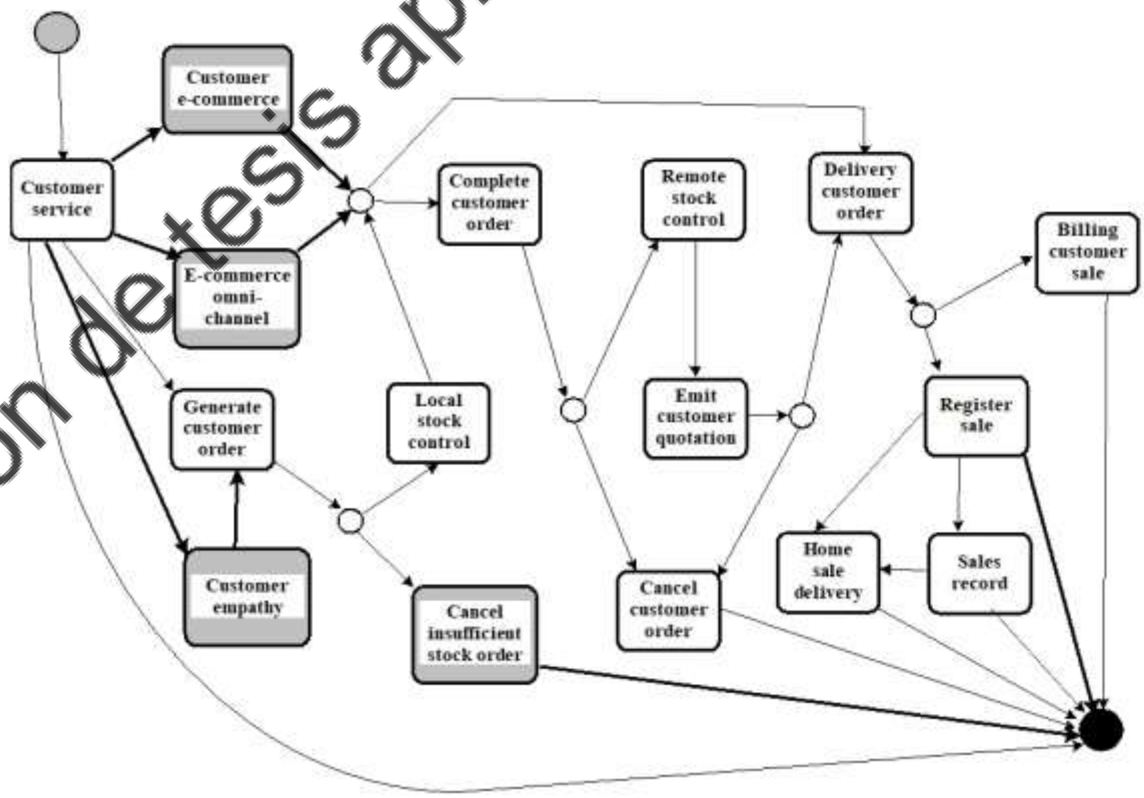


Figure 5-2. Real process model sales business.

5.4. Discussion

The quality dimensions of integrity and unambiguity are two rules that are fulfilled through the standardized design of the databases; hence, a good database design is necessary in business information systems. Compliance with the rest of the quality dimensions depends on the characteristics according to the business domain addressed.

The solution to the imperfection patterns that are treated in the present study are of general habit. This solution (manual or computerized) allows the QEL to be used, i.e., a refined event log that contains the actual processing records of the business.

Based on process mining techniques, the current process model is generated from a set of current business activities (Figure 5-1), and based on the QEL, the real business process model is generated (Figure 5-2). In addition, the application of process mining techniques [88] can verify the conformity of the real process model versus current business activities. Therefore, based on the nonconformities of the real process model (differences between the current and real process models), user strategies are identified and defined.

5.5. Conclusion

The presented method allows the QEL to be obtained for any business; provided that there is a log of events (transactions) and set of activities (manual of procedures) of the business. For the purpose of this study, we used preprocessed resources to continue with the mining process of user intentions.

CHAPTER 6

6. PROPOSED METHOD TO MINE THE USER INTENTIONS OF BUSINESS INFORMATION SYSTEMS

This chapter is based on previous publications [33], Section 6.1 provides an introduction to the intention mining method; Section 6.2 presents a review of related studies; Section 6.3 describes the materials used; Section 6.4 specifies the proposed method in detail; Section 6.5 presents experimental results; and Section 6.6 presents a discussion of the study.

6.1. Introduction

In any business, users generally carry out their daily activities by observing a procedures manual, i.e., for each business there is a current process model [86]. However, aside from current business activities, users use their own strategies in order to save time and improve results, ease of use, etc. [13]. The objective is to obtain the user strategies of business information systems and infer their intentions, for which we developed a method (Figure 6-1) that, from a knowledge base of activities for businesses in general, validates user strategies extracted from an event log of a specific sales business [33], [34].

6.2. Related works

In Table 7-1, we present a brief comparison between the aim, data, tools and results of the present study and previous research. In addition, in one of our previous studies [53], the most important advances in intention mining are reviewed.

- **Relevant Concepts.**

Process mining techniques are among data mining and machine learning techniques [19]. Modeling users' behaviors in terms of activities and ignoring underlying cognitive operators in humans, such as intentions and strategies, are the goals of process mining techniques [15]. Users' strategies can be hidden processes [21], and process mining tools can represent these processes in terms of activities [21], [74].

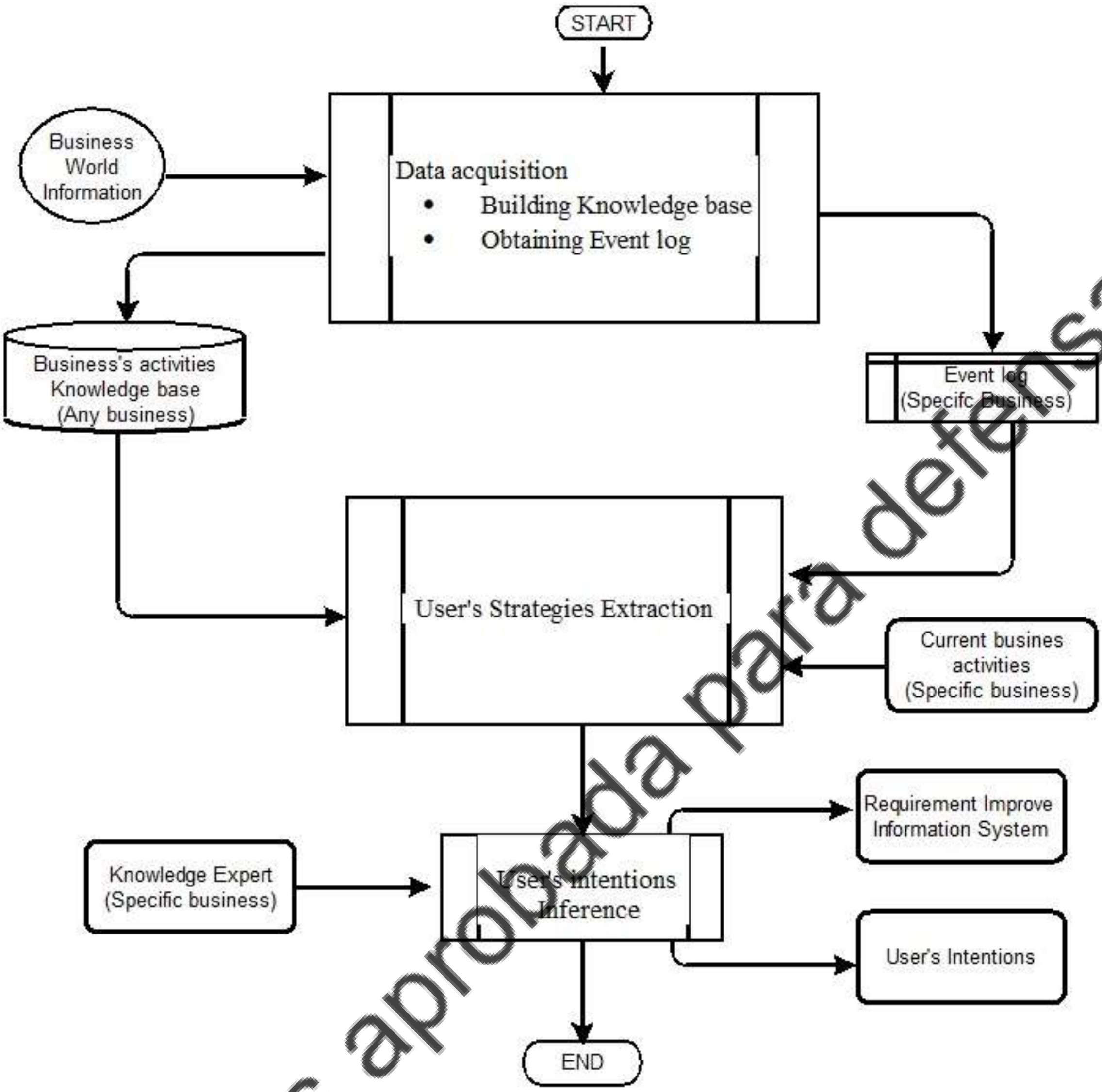


Figure 6-1. Summary of the intention mining method.

Many researchers in the field of intentional process modeling have demonstrated that the fundamental nature of processes is mostly intentional; therefore, these processes should be modeled from an intentional point of view. In this regard, nowadays, intentional process models have emerged to offer a flexible structure to model processes [15].

In some previous studies, the main objective of the use of intention mining is to extract sequences of user activities from the event logs to evaluate and predict the intentions of users regarding those activities, as can be reviewed in [76]. Process models only focus on activities, and intentional process models focus on the intentions underlying the activities rather than the activities themselves [76].

Table 6-1. Comparative data with previous works.

Previous Author	Aim	Data and Tools	Result
Present study	Discover the user intentions of business information systems	NewsIR'16 repository Knowledge base for any business Event log for specific business Business rules Knowledge expert	User intentions, possibility of user behavior inference, and identifying the improvement of information systems
[9]	Automate the construction of intentional process models	Event log Hidden Markov models	Map miner method
[15] and [19]	Focus on the construction of process models topology	Eclipse UDC processes Map miner method	Intentional process models
[74]	Design a technique to mine intentional models from traces of non-deterministic activities that follow a stochastic process	Activity traces Supervised machine learning	Intentional process models
[62]	Classifying of intentions by categories	Linguistic patterns Issue reports	Intention categories Recommendations for software developers
[89]	Analysis of intentions business functions that it can be performed effectively on short texts	Short texts (simple sentences) Machine learning	Word that express intentions
[1]	Knowledge base construction	DeepDive (databases and machine learning)	Knowledge base

A set of strategies allows users to achieve their intentions, and a certain strategy can be used to achieve several intentions [62]. The relation between intentions, strategies and activities represents the top-down structure of reasoning and acting in the cognitive processes of the human brain [15]. On the other hand, a sub-intention is associated with a parent's intention, and one intention is fulfilled if at least one of its children's sub-intentions is fulfilled [15].

Several support activities and guidance solutions based on process mining have been proposed, but its lacks suitable semantics for human reasoning and decision making and mainly relies on low-level activities [74]. Process mining aims to discover and verify the conformance of activity-oriented process models from the event log and enhance them. Intention mining has the same objectives as process mining, but it specifically addresses intentional process models [60], i.e., processes that focus on the reasoning behind the activities.

- **Intention Mining Applications.**

The map miner method (MMM) is applied to a real-world dataset, which is the event log of Eclipse UDC (Usage Data Collector) developers, [76], [76]. The resulting map process model provides a valuable understanding of the processes followed by developers and feedback on its effectiveness, as well as demonstrating the scalability of MMM.

A map specifying intentions and strategies for entity–relationship modeling was given to the students as a guide [16], [74]. In order to obtain traces, these authors developed a web-based tool that records which sections of the map were followed by students while creating an entity–relationship diagram.

To demonstrate the validity of the FlexPAISSeer tool [9] (Flex Process-Aware Information System), a single case study was conducted. The childcare system developed by 42windmills, which is used by several child daycare centers in the Netherlands, was selected as the case company, considering its suitability (the support of flexible processes through its software product).

Finally, the ProM framework is a pluggable framework that supports various plugins for different process mining techniques, such as the α-algorithm and its extensions [74].

6.3. Method

Figure 6-1 summarizes the proposed method used to extract user strategies, and based on these, infer user intentions.

6.3.1 Business Activities Mining

From news about the business world, a knowledge base (KB) is generally structured for any business. This KB contains the activities, strategies, promotions, events, etc., of the businesses that are published in the repositories [89] and kaggel. The KB was created in Postgres database management system (DBMS), and the Python language, data tables and Python code are available in the GitHub repository [90]. The steps for KB creation are specified below.

- **Business Knowledge base**

The data tables (Appendix A) of the KB and the procedures (Python code) of the method are stored and published in [90].

- **Data Pre-processing and load**

Based on traditional mining techniques, from the news articles we extracted, we transformed and loaded the text, from which we extracted the sentences that correspond to general business.

✓ **Articles extraction**

The text was extracted from 200 articles published in open access multidisciplinary business repositories. Moreover, this information was loaded into the articles data table [90]. The Python code procedure is presented in the Appendix B. The model to obtain the articles is presented.

articles = sample randomly (NewIR'16) (6-1)

✓ **Sentences extraction**

The textual content of each article was tokenized, the text was divided into sentences, and every sentence was divided into words (tokens). Tokenized and lemmatized sentences were stored in the sentences data table [90]. The Python code of this procedure is presented in Appendix C. The model to obtain the sentences is presented below.

Sentences = lemmatized (tokenized (parsing (articles))) (6-2)

✓ **NLP tag assignment (POS tag and NER tag)**

Each token in every sentence was assigned a label part of speech (POS) using Stanford's CoreNLP system [91] and the standard "Penn Treebank tagset".

We used Stanford's CoreNLP system in the tag assignment, named entity recognition (NER), according to Table 6-2. Each token was assigned a corresponding POS tag; each token, according to their POS tag, was assigned an NER tag; and the sentences tagged with POS and NER were updated. The Python code of these procedures is presented in Appendix D and Appendix E, respectively.

✓ **Strategy mentions extraction**

From the sentence data table, we extracted the tagged tokens that had sequences corresponding to the generic structures of sentences following English grammar rules.

The sequence tokens (NER-tagged) of each sentence were extracted according to the following structures: "VERB" or "NOUN" + "VERB" or "NOUN" + "VERB" + "NOUN" or "ADJECTIVE" + "NOUN" + "VERB" or "NOUN" + "VERB" + "ADVERB" or "VERB" + "ADVERB" or "ADVERB" + "VERB". Each sequence token extracted corresponds to an activity (strategy), each NER-tagged token of a strategy was replaced by their token (word), and the structured strategies were stored in the strategy_mention data table. The Python code of this procedure is presented in Appendix F.

Table 6-2. Tokens and named entity recognition.

Token Tag	**Named Entity Recognition**	**Meaning**
VB, VBD, VBG, VBN, VBP, VBZ	ACTIVITY	Verbs all modes and times
NN, NNS	NOUN	Common nouns
NNP, NNPS	PERSON/ORGANIZATION	Nouns
JJ	ADJECTIVE	Adjectives
JJR	COMPARATIVE ADJECTIVE	Comparative adjectives
JJS	SUPERLATIVE ADJECTIVES	Superlative adjectives
RB	ADVERB	Adverbs
RBR	ADVERB ADJECTIVE	Comparative adverbs
RBS	ADVERB ADJECTIVES	Superlative adverbs

- ✓ **Candidate strategies extraction**

We generated a set of strategies (S) and stored this in the candidate_strategy data table based on the strategies stored in the strategy_mention data table. For each strategy, the sentence structure was verified; duplicate strategies, punctuation marks, and nonprintable characters were eliminated; and debugged strategies were stored in the candidate_strategy data table. The Python code of this procedure is presented in Appendix G. The set of strategies is defined below:

S = {∀x / x ∈ KB and (x=”VERB” or x=“NOUN”+“VERB” or x=“NOUN”+“VERB”+“NOUN” or x=“ADJECTIVE”+“NOUN”+“VERB” or x=“NOUN”+“VERB”+“ADVERB” or x=”VERB”+”ADVERB” or x=”ADVERB”+”VERB”)} (6-3)

6.3.2 Generating Quality Event Log

In the previous chapter, the method to process and debug the GEL to generate the QEL is presented, and this QEL contains real business strategies (current business strategies plus user strategies). Based on this QEL, the real process model was generated, and based on this model, the strategies that are actually developed in a business are identified and stored in the strategy_qel data table. The Python code of this procedure is presented in Appendix H. Therefore, the QEL contains the business real process (BRP) and is defined through the following model:

BRP = {∀x / x ∈ QEL} (6-4)

6.3.3 Verifying user strategies

The real strategies of the business are obtained as a result of the verification of the BRP (specific business) against the strategies (S) contained in the candidate_strategy data table of the KB (any business). The Python code of this procedure is presented in Appendix I. The intersection of these sets of strategies yields the subset of real business strategies (BRS), as shown in the following model:

$$BRS = S \cap BRP \quad (6\text{-}5)$$

6.3.4 Validating and weighting user strategies

According to the process mining conformance rule [7], user strategies (US) are defined as the differences between the current process model and the real process model. From the current process model, the current business strategies (activities) are obtained (BCS). The Python code of this procedure is presented in Appendix J. Therefore, the US are obtained through the following model:

$$US = BRS - BCS \tag{6-6}$$

The heuristic rules of weighting are defined in two groups: a) general rules of structuring strategies for any business and b) rules for framing strategies in the domain of a specific business.

a) Strategy structuring rules (SSR):

"VERB" or "NOUN", weight = 1

"NOUN VERB" or "VERB NOUN", weight = 2

"NOUN VERB NOUN" or "ADEJCTIVE NOUN VERB" or "NOUN VERB ADVERB", weight = 3

Each of the rules with their assigned weight are defined in the following model:

$$SSR = \{\text{Catalog to any business in general}\} \tag{6-7}$$

b) Strategy of domain rule (SDR):

TRADING = ['order', 'quotation', 'stock', 'sale', 'price'].

DEALING = ['sell', 'buy', 'offer', 'promotion', 'billing', 'cancel'].

CRM = ['customer', 'empathy', 'user', 'ecommerce', 'e-commerce', 'omnichannel', 'omni-channel'].

Each of the rules has an assigned weight from 1 to 3 and are defined in the following model:

$$SDR = \{\text{Catalog in the specific business domain}\} \tag{6-8}$$

The weighting heuristic rules (WHR), are defined with the following model:

$$WHR = SSR \cup SDR \tag{6-9}$$

The US are weighted according to WHR compliance and stored in the strategy_weight data table (business Knowledge base). For this, the following model is specified.

W: weight of WHRs.

w: WHR compliance weighting by the US.

n = count (WHRs): ("n": number of weighting heuristic rules).

$m = \sum_{i=1}^{n}(W_i)$.

Therefore, the weighted user strategies (WUS) are defined by the following model.

$$WUS = \sum_{i=1}^{n}(w_i) / m \tag{6-10}$$

The Python code of this procedure is presented in Appendix K.

6.3.5 User Intentions Inference

Based on the US stored in the strategy_weight data table, the user_strategies data table is generated with US that have a weight > 0.3, as shown in Table 6-3 (According the expert on sales business [92], [95] From these strategies, it is possible infer user intentions, evaluated and supported by the participation of specific business experts [93], [94]. The obtained user intentions are presented in Table 6-4.

Table 6-3. Weighted user strategies.

	Number	Strategy name	Weighting
	1	cancel insufficient-stock order	0.75
	2	customer empathy everyday	0.75
	3	customer e-commerce	0.67
	4	customer empathy	0.67
	5	e-commerce omni-channel	0.67
	6	american news sell	0.50
	7	couple sell	0.42
	8	news sell	0.42
	9	order here	0.42
	10	order provide	0.42
	11	order spend	0.42
	12	sale register	0.42
	13	sale say	0.42
	14	school offer	0.42
	15	stock be	0.42

	16	buy	0.33
	17	cancel	0.33
	18	offer	0.33
	19	price	0.33
	20	sale	0.33
	21	sell	0.33
	22	stock	0.33

Table 6-4. User intentions.

	Number	Strategy	Intention
	1	cancel insufficient-stock order	The aim is to complete the order and not cancel it
	2	customer empathy everyday	The customer is right all the time
	3	customer e-commerce	Flexible business alternatives
	5	e-commerce omni-channel	Omnidirectional business alternatives
	7	couple sales	Complement yourself with a partner
	9	order here	Carry out the transaction here and now (time service)
	12	sale register	Business control

6.4. Results

In this case, the method is applied to the sales business. The obtained results are shown in Table 6-5.

Table 6-5. Specific sales business results

Mathematical model	Result sales business
Articles extraction articles = sample randomly (NewIR'16) (6-1)	200 articles
Sentences extraction sentences = lemmatized (tokenized (parsing (articles))) (6-2)	5501 sentences
NLP tag assignment (POS tag and NER tag) S = {∀x / x ∈ KB and (x="VERB" or x="NOUN"+"VERB" or x="NOUN"+"VERB"+"NOUN" or x="ADJECTIVE"+"NOUN"+"VERB" or x="NOUN"+"VERB"+"ADVERB" or x="VERB"+"ADVERB" or x="ADVERB"+"VERB")} (6-3)	2458 candidate strategies
Business real process (strategies) extraction BRP = {∀x / x ∈ QEL} (6-4)	billing customer sale cancel customer order cancel insufficient-stock order

	complete customer order customer e-commerce customer service customer empathy deliver customer order e-commerce omni-channel emit customer quotation generate customer order local stock control home sale delivery register sale remote stock control sales record
Business real strategies identification BRS = S ∩ BRP (6-5)	billing customer sale cancel customer order cancel insufficient-stock order complete customer order customer e-commerce customer service customer empathy deliver customer order e-commerce omni-channel emit customer quotation generate customer order local stock control home sale delivery register sale remote stock control sales record
User strategies obtaining US = BRS - BCS (6-6)	cancel insufficient-stock order customer e-commerce customer empathy e-commerce omni-channel
Strategy structuring rule SSR = {Catalog to any business in general} (6-7)	"VERB" / "NOUN" "NOUN VERB" / "VERB NOUN" "NOUN VERB NOUN / ADEJCTIVE NOUN VERB / NOUN VERB ADVERB"
Strategy of domain rule SDR = {Catalog in the specific business domain} (6-8)	TRADING = ['order', 'quotation', 'stock', 'sale', 'price'] DEALING = ['sell', 'buy', 'offer', 'promotion', 'billing', 'cancel'] CRM = ['customer', 'empathy', 'user', 'ecommerce', 'e-commerce', 'omnichannel', 'omni-channel']
Weighting heuristic rules WHR = SSR ∪ SDR (6-9)	SSR + SDR
Weighted user strategies $m = \sum_{i=1}^{n}(W_i)$.	Table 6-3.

$WUS = \sum_{i=1}^{n}(w_i)$ / m (6-10)	

6.5. Discussion

In their studies, [9] and [76] developed intention mining methods based on process mining and process-aware information systems techniques, and both authors centered their studies on recommendations of software developers. The present study is centered on the information system users.

In [62] analyzed words and sentences online to infer the user intentions of web applications. In the present study, from a knowledge base for any general business and the event log for a specific business, mining sentences (strategies) to infer the user intentions of the business information system are developed.

In [96] used a support vector machine (SVM) to classify the intentions associated with a single word that denotes intention (wish, inquiry, compare, etc.). In present study, a sentences mining method is developed based on the fundamental aspects of traditional mining with the help the NLP functions [91].

In [9] Epure et al. (2014) developed the map miner method tool to build models of intentional processes. In the present study, a method based on traditional mining techniques was developed to extract the user strategies and infer their intentions.

In [19] and [78], based on the study by [9], models of intentional processes were generated. In the present study, user strategies are formalized based on real and current business process models.

In [74], [19] and [78] generate the same models of intentional processes based on structured data (databases) (Codd, 1982). In the present study, flat news files are used, from which sentences about businesses are extracted.

In [62] and [97] derived their patterns (words and sentences) based on linguistic patterns and machine learning. In the present study, Python code was written, and NLP functions were used to extract the sentences.

In [1] built a knowledge base with the DeepDive function. In the present study, to build a knowledge base, a method of mining business sentences was developed.

Regarding the results obtained in the present study and the used tools, we conclude the following:

- NLP functions are very useful for the development of Python code;
- Only sentences with a maximum three elements can be analyzed, although there could be four or more elements in some sentences;
- The weighting heuristic rules can help extend the business domain and improve the weighting of results;
- The requirements of specific business experts could be a limitation, but also an opportunity to apply the method to any business.

CHAPTER 7

7. PROPOSED METHOD IMPLEMENTATION AND VALIDATION

Section 7.1 provides an introduction to the validation of this study; Section 7.2 presents the validation context; Section 7.3 validates the criteria that apply to this process; and finally, Section 7.4 presents the calculus precision measures.

7.1. Introduction

The validation of study's results allows the measures of suitability and applicability in the solution to the posed problem to be defined [98], [99] and [100]. In our case, we validated the results obtained from the application of the developed intention mining method.

7.2. Validation Context

The method developed in this study allows us to extract user strategies based on their inferred intentions. This method that can be applied to any business and is proven to be effective for a sales business

7.3. Criteria for Validation

From user strategies, user intentions can be inferred, in this case, with the participation of an expert in the specific sales business. Therefore, we validated the extraction of user strategies from a sales business based on the following validation criteria:

1) Binary classification:
   As we have two variables (user strategies and sales strategies), based on the method of [101] and [102], we used sensitivity to define the true positive cases and specificity to define the true negative cases; false negatives and false positives are the other two possible combinations between these two variables, which correspond to the predictions of the user strategies [102], as shown in Table 7-1.
   - True positive (TP): It is a sales strategy and a user strategy;
   - False positive (FP): It is not a sales strategy, but it is a user strategy;
   - False negative (FN): It is a sales strategy, but it is not a user strategy;
   - True negative (TN): It is not a sales strategy nor a user strategy.

Table 7-1. Prediction of the User Strategies

	User Strategy	Sales Strategy		Marginal totals
		True	False	
Prediction	True	True Positive	False Positive	TP + FP
	False	False Negative	True Negative	FN + TN
	Marginal Totals	TP + FN	FP + TN	

2) Precision measures

- Current sales strategies that are user strategies (sensitivity):

  Sensitivity = TP / (TP+FP) (7-1)

- User strategies that are not current sales strategies (specificity):

  Specificity = TN / (FN+TN) (7-2)

- Probability that user strategies are not current sales strategies (positive predicted value: PPV):

  PPV = FP / (TP+FN) (7-3)

- Probability that strategies are not current sales strategies or a user strategy (negative predicted value: NPV):

  NPV = FN / (FN+TN) (7-4)

- Probability the method predicts the presence or absence of user strategies (Accuracy):

  Accuracy = (FP + FN) / (TP + FP + FN + TN) (7-5)

7.4. Validation

According to the results obtained from our method (Table 6-3), and based on the current and real process models of the sales business, the results obtained for each variable in question (user strategies and sales strategies) were quantified:

TP: 1 Current sales strategies that are also user strategies;

FP: 13 User strategies that are not current sales strategies;

FN: 0 Current sales strategies that are not user strategies;

TN: 8 Strategies that are not current sales strategies or user strategies.

The computing of the precision measures are presented below, and the algorithm of this procedure is presented in Appendix L.

Sensitivity = 0,07142857

Specificity = 1

Positive Predicted Value = 0,92857143

Negative Predicted Value = 0

Accuracy = 0,59090909

The sensitivity is quite low, which means there is a low proportion of true positives, i.e., the cases of sales strategies that are also user strategies are scarce.

Due to a low sensitivity, a high specificity is expected [102]; therefore, the rate of false negatives should be low, which means that, the lower the number of extracted current sales strategies, the higher the probability of user strategies.

For the positive predicted value, the objective is to predict user strategies that are not current sales strategies; therefore, a significant prediction value is expected.

For the negative predicted value, the value should be as low as possible. In every instance, it seeks to extract sales strategies, with user strategies taking priority, but if different strategies are extracted, it is expected that they will be in minimal proportions.

Finally, an acceptable precision value was obtained; this could suggest that the method is acceptable, with 68% effectiveness.

CHAPTER 8

8. RESULT DISCUSSION

In order to achieve the stated objective, based on the user strategies extracted from the event log, the user intentions of the specific business information system (sales) were identified. The objective was achieved through the activities specified in Figure 1-1. The results obtained are discussed below.

8.1. Non-feasible Scenario

In principle, the issue of applying intention mining to the governmental organization "Instituto Geofísico - Escuela Politécnica Nacional" (IG-EPN) is addressed, which has an information system that can generate an abundant event log of seismological activities. After the feasibility study, it was determined that this scenario, does not have the necessary elements that characterize a business, where the users are the main resource used to reach the objective of this study: "Intention mining of an information system user". However, to arrive at this conclusion, a current situation survey of this scenario was developed, which inspired the study by [32]. The study was rewritten with a focus on business scenarios in [53], where a literature review was developed and state-of-the-art research was discussed. Based on this research, this present study on intention mining was developed, which centers on the user and is markedly different from previous research, where the aims were to give recommendations to software developers.

8.2. The State of the Art

Between 2010 and 2018, the development of the topic "intention mining" was identified in literature reviews in the context of business information systems, based on which this study was developed. In addition, research published between 2019 and 2021, and other relevant studies in this context, are identified and analyzed below.

[70]: "From our analysis, the challenges of user intent mining fall into three folds. Firstly, user intent could be express explicitly or implicitly. Implicit user intents do not contain the intent keywords, which is more challenging to classify and recognize users' real ideas. Secondly, research of user intent in many domains is lacking. Thirdly, we also observed that user intent is not stable but changing over time. Intentions could interact with each other and have a time decaying phenomenon. Then how to model this dynamic nature of intention is also important to predict user's interests and information needs".

This author addresses the explicit intentions underlying user strategies. I agree with this author. The research of user intent in many domains is lacking, and the research domain of the present study is business information systems.

[14]: “The user intention mining with respect to business perspective is an important and challenging task due to the varying nature of customer-generated text data. The purpose of this review is to present a brief review of studies pertaining to user intention mining with emphasis on discussing different machine learning and deep learning techniques”.

These authors develop a literature review to extract the intentions of business users, using machine learning and deep learning tools applied to commercial transactions and performed by customers through social networks. Our study focuses on the intentions of the user in a business information system.

[62]: “In this paper, we manually categorize 5,408 sentences from issue reports of four projects in GitHub. We propose a deep learning based approach to automatically and more accurately classify sentences into different categories of intentions. A case study on four open source projects with 2,076 issue reports shows that our approach achieves an average of the 68.7%”.

In contrast, these authors classify sentences (possible intentions) manually and achieve an average of 68.7% performance. In the present study, sentences are extracted from news articles through a proprietary method of sentence mining using NLP libraries [91]. Additionally, a knowledge base is created for general businesses, which serves to verify and validate user strategies (intentions); in the best case scenario, the obtained results suggest a performance of 75%.

[71]: “Today vast and diverse event records of applications exist for almost every scientific domain, making their integration and intelligent exploitation challenging. Intention mining is the ability to predict a user’s goals. Knowing the user’s intention can support the decision-making of the network administrators. The main input of all algorithm used to discover intentional process model is the log file (traces activities), which is unstructured dataset and not ready to be feed as-is to machine learning algorithm. Therefore, this paper aims to describe the data preprocessing steps, which transform the unstructured log file to a structured one”.

Similar to these authors, in the present study, the main resource is the event log, which is preprocessed and converted into structured data (quality event log). In addition, user intentions are identified, allowing us to understand their behavior, which facilitates decision making for business managers.

In [72]: "The aim of this work is to conduct a literature review about Intents, Intention Mining and Intent Classification. Nowadays, Intention Mining is widely used in the Information Systems Engineering field. This paper mainly focuses and discusses on the literature review algorithms, models and tools used in Intention Mining. We hope that this information will be useful for developing models to retrieve intentions from the traces of activities and developing various intention mining techniques, which will allow identifying the gaps between the prescribed processes and the actual processes of a business".

Essentially, this author reviews the literature on "mining of intentions", which is one aspect of the unique literature review developed in this study.

In [73]: "Users use the network more and more frequently, and more and more data is published on the network. Therefore, how to find, organize, and use the useful information behind these massive data through selective means, and analyze user intentions is a huge challenge".

This author develops a technique for mining the intentions of users of data networks based on temporal networks, improving an already known method. It is not related to the present study.

8.3. Why Database Design as Quality Dimension?

The event log is generated by default in a business information system due to the processing of data stored in the databases. The quality of the information obtained from an information system depends on the quality of the processed data and the design of structures (databases) where they are stored for easy, timely and appropriate access [5]. Consequently, the information quality contained in the event log depends on the high-quality design of databases.

The generic event log (GEL) is generally used for auditing processes through the application of process mining to determine shortcomings in the execution of business activities and, especially, to identify fraud. In the present study, this GEL is the source of the extracted user strategies. The GEL in this case is structured based on the Sales_Log, Sales_Principles, Sales_Policies and Sales_Rules data files published in the GitHub repository. These files were obtained from the forums Kaggel, SignalMedia, DBpedia, BigQuery, etc.

8.4. Why a Quality Event Log?

The GEL that is generated by the information system generally contains irrelevant data for the present study, and the data of interest contain errors and imperfection patterns, which are corrected through the method specified in [20]. As a result, a quality event log (QEL) is obtained [33], from which user strategies can be extracted, and their user intentions can be inferred. Practically, the generation of this GEL corresponds to the preprocessing of a resource to be mined using a traditional mining tool.

8.5. User Strategies Formalization

On the one hand, a knowledge base is created for any general business, which includes business activities and strategies. On the other hand, you can obtain an event log of the specific sales business:

- The existence of the activities and strategies of the event log are verified in the knowledge base and the verified strategies are extracted;
- By applying process mining [35] to the verified strategies, the actual process model of the sales business is generated (Figure 7-2);
- By applying process mining, the current process model is generated according to the activities of a sales business [20].
- User strategies are extracted from the differences between the actual process model and the current process model.
- The strategies are validated and weighted, according to the formal structure of English sentences and the fulfillment of the heuristic rules of the sales process.
- The strategies with a weighting higher than 30% are extracted and based on these, user intentions are inferred.

To date, there are no authors who have explored this process; this study is the first to research this topic [34].

8.6. User Intentions Inference

According to the degree of instantiation (verification and validation) of the characteristics between the strategies extracted from the knowledge base and the strategies obtained from the event log of the sales business, user strategies were formalized. These weighted user strategies and the criterion of an expert in the sales business have made it possible to infer user intentions in business information system [34].

This is the final product, with which the objective of this study on the novel subject of "intention mining" has been achieved.

CHAPTER 9

9. SUMMARY

9.1. Conclusions

- **For Any Business**

According to the resilience of each business and the feasibility of having a log of activities (events), it is possible to innovate information systems through the knowledge of user behavior. This can be achieved if their intentions in the development of the daily activities can be determined. Although in the development of this study, results have been obtained for the specific business of sales, mathematical models (Table 7-6) have been developed to generalize the process and apply it to any business, provided that there is a log of events for that business.

- **Research Novel Subject**

The subject of intention mining in the context of information systems has not been developed on a large scale until the present. Among the literature reviews published between 2010 and 2018, [19] highlights this topic in the context of information systems. From 2019 to the present day, the development of this topic has been well-received in the field of social networks, but for the field of information systems, it is still scarce. In the "Result Discussion" section, several publications [14], [62], [70]-[73] are discussed.

- **Aims**

The research objective was met to the extent that the hypotheses have been tested and the research questions have been answered.

RQ1. In a business, are the current processes, actual processes and functions of the person who executes them known and promulgated?

From the execution of the activities based on the current business process, the generic event log (GEL) is obtained; this log corresponds to the real activities of the business; that is, in this log the activities are registered according to the current business process model; in addition to user strategies. The points that answer the RQ1 are specified below.

- ✓ Current business process model, Figure 7-3.
- ✓ Generic event log (GEL), Table 5-3.
- ✓ Real business process model, Figure 7-2.
- ✓ User strategies, Table 7-3.

RQ2. In a business information system, is there a relationship between the current process, the real process and the users who execute them?

Through the application of process mining, the differences between the current process models and the real business process are identified; Based on these differences, the user's strategies are extracted and debugged. With which the answer to RQ2 is given

- ✓ Differences between real and current business process models are shown in Figure 7-2 and Figure 7-3.
- ✓ Debugged user strategies are stored in GitHub repository [90] and shown in Appendix 7-G.

RQ3. How to infer user intentions?

Finally, in order to respond to the RQ3 main objective of this research work, the steps of the method proposed in chapter 7 are specified below; which in summary consists: The creation of a knowledge base of the business in general, the user strategies contained in the quality event log (QEL), the verification of the user strategies in the knowledge base, the validation and weighting of the user's strategies based on compliance with the rules of the specific sales business and, the inference of the user's intentions based on their strategies.

- ✓ A knowledge base created with our own method and stored in the GitHub repository [90] and shown in Appendix 7-A to Appendix 7-F.
- ✓ User strategies: Table 7-4 and Appendix 7-H, Appendix 7-I
- ✓ Validated and weighted user strategies: Appendix 7-J.
- ✓ Inferred user intentions: Table 7-5.

- **Limitations and Scope**

Initially, the development of a solution for a specific business was envisaged, but as the study progressed and free data became available, the application was generalized to any business based on the mathematical models specified in Table 7-6. However, the limitation of this study lies in its specificity to the context of business information systems.

- **Contributions**

In terms of fulfilling the proposed objectives and responding to the research questions by satisfying the hypotheses raised, the following are the relevant contributions:

✓ Literature review method for intention mining in the field of information systems [53].

✓ Methodology for designing relational databases based on a scenario analysis of their policies and business rules [5].

✓ Method to generate a quality event log from a generic event log [32], [20].

✓ A knowledge-base-building method through a sentence mining method [33].

✓ The inference of user intentions through an intention mining method [34].

9.2. Thesis Summary

In this study, prior to addressing the business scenario in the context of information systems, the first steps were taken in data mining (seismological events), and the scenario for a study on this new topic was defined: "intent mining". Following the guidelines of traditional mining, methods have been developed, making it possible to achieve the objective of inferring user intentions in a business information system.

Our method of literature review has helped to define state-of-the-art research and understand the current situation regarding the development of "intention mining". The method used to generate the quality event log (indispensable for the mining process), has made it possible to identify and define business activities (current process model), in addition to identifying user strategies (differences between the current and real business process models), based on which user intentions can be inferred. In addition, a method of mining sentences (activities) from the business world was developed, in order to build a knowledge base for businesses in general, which can verify the user strategies of a specific sales business.

9.3. Future Studies

In future studies, we propose developing a method that can identify the gap between the current and actual processes of a business. For this, it is necessary to know the state-of-the-art advances in intention mining. The first step for future research is to develop this method of literature review, which involves concepts such as: event log, event, process instance, process traces, user strategies, user intentions, process mining, and intention mining, etc. (described in the glossary). Topics for future research are suggested below:

- Implement the applications of each of the proposed methods;
- Develop applications to generate business recommendations that aim to improve user-centered processes;
- Improve the research developed through the implementation of an unsupervised/supervised machine learning algorithm;
- Promote flexible processes through the development of process-aware information systems, based on the intention mining techniques;
- Develop an intelligent system to provide recommendations for improving business information systems based on user strategies and intentions.

9.4. List of Publications

- **JOURNAL ARTICLES**

Title: Strategy Mining for Inferring Business Information System User Intentions

Journal: Applied Science (Switzerland)

Ranking: JCR Q2, SJR Q3

Authors: Oswaldo Díaz and Maria Pérez

Date Publication: 2022/06/11

DOI: doi.org/10.3390/app12125949

Abstract:

The aim of this study was to identify user strategies to infer their intentions in developing activities, from the current process model and the real process model of the business. A user's intentions can be used to identify their behavior, and to define the requirements for improving the business information system. The presented method follows the guidelines of the current mining tools, and it is supported by a knowledge base of businesses in general and the event log for a specific business. The user strategies are validated and weighted through the rules of the specific business. The user intentions are inferred based on their strategies and the knowledge of an expert within the specific business. The method is applied to a specific sales business, and the obtained results suggest that the proposed method can extract 75% of user intentions. In addition, the method is generalized to apply it to any business, as long as we can obtain the event log and the rules of the business

Relationship with thesis:

This paper contains the method of mining sentences (activities) extracted from news articles from the business world. It also contains the method of formalizing the user's strategies and the inference of the user's intentions of the business information system. The paper is the base of chapters 6 and 7 of this thesis

Title: Intention Mining from Knowledge base and Supervised Machine Learning

Journal: 3Ciencias TIC

Ranking: ESCI

Authors: Oswaldo Díaz and Maria Pérez

Date Publication: 2021/09/29

DOI: doi.org/10.17993/3ctic.2021.103.65-101

Abstract:

The lack of flexibility in information systems has led users to use their own strategies to carry out their daily activities in line with business objectives. In this way, users fulfill their functions, improve their performance and save resources; especially time. An intent mining method is proposed, which is based on supervised machine learning, supported by a knowledge base and heuristic business rules. A knowledge base is structured from multidisciplinary business documents for any business in general. From the application of the proposed method to a particular business (sales), the user's strategies are obtained and from these, their intentions in the development of their activities in the sales business are inferred, through an information system. This work suggests the development of flexible information systems and provides business managers with a tool to identify and implement new business strategies, based on user strategies.

Relationship with thesis:

An alternative method to the method presented in Chapter 6 is presented in this publication.

Title: Literature Review about Intention Mining in Information Systems

Journal: Journal of Computing Information System

Ranking: JCR Q3, SJR Q2

Authors: Oswaldo Díaz, Maria Pérez and Jorge Lascano

Date Publication: 2019/06/16

DOI: 10.1080/08874417.2019.1633569

Abstract:

Based on the event log, the process mining techniques allow the discovery of the real processes, the verification of conformity with the prescribed processes and the improvement them. Also, in the current literature has evidenced the development of some works of process mining to the discovery of user intentions (intention mining). The purpose of this work is to conduct a literature review about the development of intention mining in the information systems engineering area, which aims is to define the state of

the art of the intention mining techniques that are have developed based on the process mining techniques. Then, is developed the literature review, it analyzes and discusses the identified resources in the review. Is estimated that these resources will be the starting point to develop an intention mining technique, which allows identifying the gap between the prescribed processes and the real processes of a business.

Relationship with thesis:

This publication is the support of chapters 3 and 4 of the thesis; literature review and state of the art.

Title: Methodology for Designing Relational Databases Based on Scenario Analysis their

Policies and Business Rules

Journal: 3Ciencias TIC

Ranking: ESCI

Authors: Oswaldo Díaz

Date Publication: 2015/09/19

DOI: dx.doi.org/10.17993/3ctic.2015.43

Abstract:

The critical success factor in database design for a business is that; the database gives facilities to meet the information requirements according to business rules and the scenario's policies; the relational model is a reference for the methodology proposed in this paper and the relational algebra (atomicity; dependence and transitivity) is used only in the normalization; the methodology is based on the scenario analysis; the conceptualization of business through its rules and experience of the designer; the methodology was applied to a case study and obtained the design of the database in third normal form.

Relationship with thesis:

Database design is a dimension of quality in the generation of the quality event log from the generic event log, produced by the information system, through the method presented in chapter 5 of the thesis.

- **CONFERENCE ARTICLES:**

Title: Quality Event Log to Intention Mining: A Study Case.

Proceeding: 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA).

Authors: Oswaldo Díaz y María Pérez

Date Publication: 2020/03/13

DOI: 10.1109/ICCSEA49143.2020.9132856

Abstract:

In a business, the information system produces a generic event log (GEL), which contains information on daily activities, usually, this GEL is too large and its content isn't of quality. In the present work a method based on process mining is proposed, to generate a quality event log (QEL) and whose content would allow to discover the user's strategies (user's intentions), that in a future work would be formalized and modeled through any method of intention mining. In addition, is presented a study case about a process of sales business.

Relationship with thesis:

In this publication, a case study is presented, as a preliminary step for the method of generating the quality event log of chapter 5

Title: Log Design for Storing Seismic Event Characteristics Using Process, Text, and Opinion Mining Techniques

Proceeding: 2018 International Conference on eDemocracy & eGovernment (ICEDEG)

Authors: Oswaldo Díaz y María Pérez

Date Publication: 2018/04/04

DOI: 10.1109/ICEDEG.2018.8372312

Abstract:

A fundamental preliminary element for discovering knowledge is the event log. This discovery knowledge from the log can help us to analyze and recognize certain patterns (features) that occur during seismic events. Therefore, in this work, we propose to design the log for seismic events from the information stored in the MySQL database, through the SeisComp3 system and the information of the scientific-technical staff of Instituto Geofísico at Escuela Politécnica Nacional. The design of the mentioned log is the product of the integration of the structures of the models obtained from text, process and opinion mining applied to the primary data with the log-book and the reports of the experts in seismology available in this Institute. Thus, the log will store data that would allow identifying the features of these physical phenomena through intention mining techniques.

Relationship with thesis:

This publication is the product of a research project, which served to define the feasibility of the scenario (for the development of the research topic) specified in chapter 2

References

[1] C. De Sa *et al.*, “Incremental knowledge base construction using DeepDive,” *VLDB J.*, vol. 26, no. 1, pp. 81–105, 2017, doi: 10.1007/s00778-016-0437-2.

[2] J. C. S. Do Prado Leite and M. C. Leonardi, “Business rules as organizational policies,” *Proc. 9th Int. Work. Softw. Specif. Des. IWSSD 1998*, no. June 2014, pp. 68–76, 1998, doi: 10.1109/iwssd.1998.667921.

[3] P. C. y J. V. H. Almanza Rebeca, “La resiliencia empresarial elemento clave en el cambio organizacional,” *Rev. Iberoam. Gob. Local*, vol. 10, no. January 2016, p. 16, 2017, [Online]. Available: https://www.researchgate.net/profile/Jose_Vargas-Hernandez/publication/314279800_La_resiliencia_empresarial_elemento_clave_en_el_cambio_organizacional/links/58beed6492851cd83aa12358/La-resiliencia-empresarial-elemento-clave-en-el-cambio-organizacional.pdf

[4] N. Arshed and M. Danson, “Enterprise Policy,” *Enterp. Concepts Issues*, no. January, 2016, doi: 10.23912/978-1-910158-75-3-2882.

[5] O. Díaz Rodríguez, “Methodology for Designing Relational Databases Based on Scenario Analysis Their Policies and Business Rules,” *TIC*, vol. 344, no. 9, pp. 77–88, 2015, [Online]. Available: https://www.3ciencias.com/wp-content/uploads/2015/09/METODOLOGÍA-PARA-DISEÑAR-BASES-DE-DATOS-RELACIONALES-CON-BASE-EN-EL-ANÁLISIS-DE-ESCENARIOS-SUS-POLÍTICAS-Y-LAS-REGLAS-DEL-NEGOCIO.pdf

[6] Kimball Group, “A Dimensional Modeling Manifesto - Kimball Group.” https://www.kimballgroup.com/1997/08/a-dimensional-modeling-manifesto/ (accessed Oct. 18, 2022).

[7] W. M. P. Van Der Aalst *et al.*, “Manifiesto sobre Minería de Procesos. Traducido desde The Process Mining Manifesto by the IEEE Task Force on Process Mining,” *BPM 2011 Work. proceedings, Lect. Notes Bus. Inf. Process.*, 2011, [Online]. Available: http://www.win.tue.nl/ieeetfpm

[8] M. Ghasemi and D. Amyot, “From event logs to goals: a systematic literature review of goal-oriented process mining,” *Requir. Eng.*, vol. 25, no. 1, pp. 67–93, 2020, doi: 10.1007/s00766-018-00308-3.

[9] E. V. Epure, C. Hug, R. Deneckére, and S. Brinkkemper, “What shall I do next? Intention mining for flexible process enactment,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8484 LNCS, pp. 473–487, 2014, doi: 10.1007/978-3-319-07881-6_32.

[10] Merriam-Webster, “Event Definition & Meaning - Merriam-Webster.” https://www.merriam-webster.com/dictionary/event (accessed Oct. 18, 2022).

[11] Generic Event Log, “Generic Windows Event Log | InsightIDR Documentation.” https://docs.rapid7.com/insightidr/generic-windows-event-log/ (accessed Oct. 18, 2022).

[12] G. S. Do Nascimento, C. Iochpe, L. H. Thom, and M. Reichert, “A method for rewriting legacy systems using business process management technology,” *ICEIS 2009 - 11th Int. Conf. Enterp. Inf. Syst. Proc.*, vol. ISAS, no. December 2019, pp. 57–62, 2009, doi: 10.5220/0001950400570062.

[13] A. D. Lloyd, R. Dewar, and R. Pooley, “Legacy Information Systems and Business Process Change: A Patterns Perspective,” *Commun. Assoc. Inf. Syst.*, vol. 2, no. December, 1999, doi: 10.17705/1cais.00224.

[14] A. Habib, F. K. Saddozai, A. Sattar, A. Khan, I. A. Hameed, and F. M. Kundi, “User Intention Mining in Bussiness Reviews: A Review,” *Proc. - 2018 5th Int. Conf. Behav. Econ. Socio-Cultural Comput. BESC 2018*, pp. 243–249, 2018, doi: 10.1109/BESC.2018.8697303.

[15] G. Khodabandelou, C. Hug, and C. Salinesi, “A novel approach to process mining: Intentional process models discovery,” *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, 2014, doi: 10.1109/RCIS.2014.6861040.

[16] W. Van der Aalst, " Process Mining: The Missing Link," in *Process mining: Data*

science in action, Seccond Edition Springer 2016. pp: 35-37.

[17] D. A. Fischer, K. Goel, R. Andrews, C. G. J. van Dun, M. T. Wynn, and M. Röglinger, "Enhancing event log quality: Detecting and quantifying timestamp imperfections," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12168 LNCS, pp. 309–326, 2020, doi: 10.1007/978-3-030-58666-9_18.

[18] E. Brzychczy and A. Trzcionkowska, *Creation of an event log from a low-level machinery monitoring system for process mining purposes*, vol. 11315 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-030-03496-2_7.

[19] G. Khodabandelou, "Contextual recommendations using intention mining on process traces: Doctoral consortium paper," *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, 2013, doi: 10.1109/RCIS.2013.6577728.

[20] O. E. Díaz-Rodriguez and M. Gabriela Perez Hernandez, "Quality Event Log to Intention Mining: A Study Case," *2020 Int. Conf. Comput. Sci. Eng. Appl. ICCSEA 2020*, 2020, doi: 10.1109/ICCSEA49143.2020.9132856.

[21] G. Khodabandelou, C. Hug, and C. Salinesi, *Mining users' intents from logs*, vol. 6, no. 2. 2015. doi: 10.4018/IJISMD.2015040102.

[22] M. Sumner and R. Kleppe, "Information Systems Strateg Y End-Use R a N D Applicatio N Developmen T," *ACM SIGMIS Database DATABASE Adv. Inf. Syst.*, no. February 1983, pp. 19–30, 1987, [Online]. Available: http://delivery.acm.org/10.1145/1020000/1017818/p19-sumner.pdf?ip=137.122.64.159&id=1017818&acc=ACTIVE SERVICE&key=FD0067F557510FFB.8B52749AC303A1AA.4D4702B0C3E38B35.4D4702B0C3E38B35&__acm__=1545184255_918d9a2959aa007e458070804306fb37

[23] K. Olson, "Process Mining Concepts for Discovering User Behavioral Patterns in Instrumented Software," *Undefined*, 2017, [Online]. Available:

https://www.semanticscholar.org/paper/Process-Mining-Concepts-for-Discovering-User-in-Olson/12fffc4514ca5927fc4cc224784c69374861d863

[24] M. Jans, “Building a valuable event log for process mining : an experimental exploration of a guided process Building a valuable event log for process mining : an experimental exploration of a guided process,” no. March, 2019, doi: 10.1080/17517575.2019.1587788.

[25] C. Monsalve, A. April, and A. Abran, “Business Process Modeling with Levels of Abstraction A Survey of Practioners with Experience,” *Ieee Colcom*, pp. 1–6, 2015, [Online]. Available: http://ieeexplore.ieee.org/document/7152080/

[26] L. Floridi, “The method of levels of abstraction,” *Minds Mach.*, vol. 18, no. 3, pp. 303–329, 2008, doi: 10.1007/s11023-008-9113-7.

[27] Ł. Budnik and H. Krawczyk, “Dynamic analysis of enterprise business scenarios,” *Proc. - IEEE Int. Enterp. Distrib. Object Comput. Work. EDOC*, pp. 112–121, 2011, doi: 10.1109/EDOCW.2011.40.

[28] E. F. Codd, “A Relational Model of Data for Large Shared Data Banks,” *Commun. ACM*, vol. 26, no. 1, pp. 64–69, 1983, doi: 10.1145/357980.358007.

[29] F. L. Gallego, “Administración, organización y empresa,” vol. 41, pp. 9–18, 2005, [Online]. Available: https://www.redalyc.org/pdf/215/21513702.pdf

[30] A. R. Qastharin and D. Liu, “Business Model Differences Between Business and Social Enterprise,” pp. 1–22, 2014, [Online]. Available: https://www.academia.edu/29840740/Business_Model_Differences_Between_Business_and_Social_Enterprise

[31] E. Guba and Y. Lincoln, “Paradigmas en competencia en la investigación educativa,” *Por los rincones. Antología métodos Cual. en la Investig. Soc.*, pp. 113–145, 2002.

[32] O. E. Diaz-Rodriguez and M. Perez, “Log Design for Storing Seismic Event Characteristics Using Process, Text, and Opinion Mining Techniques,” *2018 5th Int.*

Conf. eDemocracy eGovernment, ICEDEG 2018, pp. 281–285, 2018, doi: 10.1109/ICEDEG.2018.8372312.

[33] O. E. Díaz-Rodríguez and M. G. Pérez Hernández, “Minería de intenciones a partir de una base del conocimiento y aprendizaje automático supervisado,” *3C TIC Cuad. Desarro. Apl. a las TIC*, vol. 10, no. 3, pp. 65–101, 2021, doi: 10.17993/3ctic.2021.103.65-101.

[34] O. Diaz and P. Maria, “Strategy Mining for Inferring Business Information System User Intentions,” *Appl. Sience*, 2022.

[35] X. Xia, W. Song, F. Chen, X. Li, and P. Zhang, “Effa: A ProM plugin for recovering event logs,” *ACM Int. Conf. Proceeding Ser.*, vol. 18-Septemb, pp. 108–111, 2016, doi: 10.1145/2993717.2993732.

[36] A. Ratner and C. Ré, “Research for practice: Knowledge base construction in the Machine-Learning Era,” *Commun. ACM*, vol. 61, no. 11, pp. 95–97, 2018, doi: 10.1145/3233243.

[37] I. Flores, H. Núñez, E. Ramos, J. Vitelli, and V. Rodríguez, “Construcción de bases de conocimiento en el dominio de la microbiología mediante técnicas de aprendizaje automático Construcción de bases de conocimiento en el dominio de la microbiología mediante técnicas de aprendizaje automático,” no. September, 2010.

[38] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A design science research methodology for information systems research,” *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.

[39] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, “Design science research evaluation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7286 LNCS, pp. 398–410, 2012, doi: 10.1007/978-3-642-29863-9_29.

[40] J. Rosney, “I El Macroscopio (Rosnay) - EL MACROSCOPIO JOEL DE ROSNAY Capítulo 1. 1.- A través del macroscopio. - StuDocu.” https://www.studocu.com/en-

us/document/universidad-bolivariana/salud/i-el-macroscopio-rosnay/2953618?origin=home-recent-1 (accessed Oct. 18, 2022).

[41] R. J. Qureshi and D. Road, “Requirements and the baseline plan,” no. May, pp. 1–4, 2012, [Online]. Available: http://arxiv.org/pdf/1201.4500.pdf

[42] A. Rozinat and W. M. P. van der Aalst, “Conformance checking of processes based on monitoring real behavior,” *Inf. Syst.*, vol. 33, no. 1, pp. 64–95, 2008, doi: 10.1016/j.is.2007.07.001.

[43] A. Berisha - Shaqiri, “Management Information System and Decision-Making,” *Acad. J. Interdiscip. Stud.*, no. December, 2014, doi: 10.5901/ajis.2014.v3n2p19.

[44] M. Alidoost Nia and A. Ruiz-Martínez, “Systematic literature review on the state of the art and future research work in anonymous communications systems,” *Comput. Electr. Eng.*, vol. 69, no. December, pp. 497–520, 2018, doi: 10.1016/j.compeleceng.2017.11.027.

[45] SeisComP3, “Seiscomp3 Development.” https://www.seiscomp.de/seiscomp3/ (accessed Oct. 18, 2022).

[46] QuakeML, “QuakeML - QuakeML.” https://quake.ethz.ch/quakeml (accessed Oct. 18, 2022).

[47] IRIS, “IRIS: Seismic Event Data.” http://ds.iris.edu/ds/nodes/dmc/data/types/events/ (accessed Oct. 18, 2022).

[48] T. K. Ahern, R. Casey, D. Barnes, R. Benson, and T. Knight, “Seed Reference Manual,” *Int. Database Syst. Philadelphia, Pa*, p. 224, 2014, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:SEED+Reference+Manual#4

[49] W. Jin, H. H. Ho, and R. K. Srihari, “OpinionMiner: A novel machine learning system for web opinion mining and extraction,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1195–1203, 2009, doi: 10.1145/1557019.1557148.

[50] University of Virginia Library, “Reading PDF files into R for text mining | University

of Virginia Library Research Data Services + Sciences.” https://data.library.virginia.edu/reading-pdf-files-into-r-for-text-mining/ (accessed Oct. 18, 2022).

[51] S. Dang and P. H. Ahmad, “Text Mining : Techniques and its Application,” *IJETI Int. J. Eng. Technol. Innov.*, vol. 1, no. 4, pp. 22–25, 2014, [Online]. Available: www.ijeti.com

[52] S. VijayGaikwad, A. Chaugule, and P. Patil, “Text Mining Methods and Techniques,” *Int. J. Comput. Appl.*, vol. 85, no. 17, pp. 42–45, 2014, doi: 10.5120/14937-3507.

[53] O. E. Diaz, M. G. Perez, and J. E. Lascano, “Literature Review about Intention Mining in Information Systems,” *J. Comput. Inf. Syst.*, vol. 0, no. 0, pp. 1–10, 2019, doi: 10.1080/08874417.2019.1633569.

[54] A. Goldstein, T. Johanndeiter, and U. Frank, *Business process runtime models: towards bridging the gap between design, enactment, and evaluation of business processes*, vol. 17, no. 1. Springer Berlin Heidelberg, 2019. doi: 10.1007/s10257-018-0374-2.

[55] C. Okoli and K. Schabram, “Working Papers on Information Systems A Guide to Conducting a Systematic Literature Review of Information Systems Research,” *Work. Pap. Inf. Syst.*, vol. 10, no. 2010, 2010, doi: 10.2139/ssrn.1954824.

[56] Y. Levy and T. J. Ellis, “A systems approach to conduct an effective literature review in support of information systems research,” *Informing Sci.*, vol. 9, no. May 2014, pp. 181–211, 2006, doi: 10.28945/479.

[57] A. Amrollahi, A. H. Ghapanchi, and A. T.-K. Talaei-Khoei, “A Systematic Literature Review on Strategic Information Systems Planning: Insights from the Past Decade,” *Pacific Asia J. Assoc. Inf. Syst.*, vol. 5, no. 2, pp. 39–66, 2013, doi: 10.17705/1pais.05203.

[58] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S.

Linkman, “Systematic literature reviews in software engineering - A systematic literature review,” *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.

[59] W. Bandara, S. Miskon, and E. Fielt, “A systematic, tool-supported method for conducting literature reviews in information systems,” *19th Eur. Conf. Inf. Syst. ECIS 2011*, 2011.

[60] E. Epure, C. Hug, R. Deneckère, and S. Brinkkemper, “Intention-mining: A solution to process participant support in process aware information systems,” *Dep. Inf. Comput. Sci. Utr. Univ. Utrecht, Netherlands*, p. 120, 2013, [Online]. Available: http://www.cs.uu.nl/research/techreps/repo/CS-2013/2013-020.pdf

[61] E. V. Epure, R. Deneckere, and C. Salinesi, “Analyzing perceived intentions of public health-related communication on Twitter,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10259 LNAI, pp. 182–192, 2017, doi: 10.1007/978-3-319-59758-4_19.

[62] Q. Huang, X. Xia, D. Lo, and G. C. Murphy, “Automating Intention Mining,” *IEEE Trans. Softw. Eng.*, vol. PP, no. c, p. 1, 2018, doi: 10.1109/TSE.2018.2876340.

[63] A. Di Sorbo, S. Panichella, C. A. Visaggio, M. Di Penta, G. Canfora, and H. C. Gall, “Development emails content analyzer: Intention mining in developer discussions,” *Proc. - 2015 30th IEEE/ACM Int. Conf. Autom. Softw. Eng. ASE 2015*, pp. 12–23, 2016, doi: 10.1109/ASE.2015.12.

[64] N. Banerjee, D. Chakraborty, A. Joshi, S. Mittal, A. Rai, and B. Ravindran, “Towards analyzing micro-blogs for detection and classification of real-time intentions,” *ICWSM 2012 - Proc. 6th Int. AAAI Conf. Weblogs Soc. Media*, no. January, pp. 391–394, 2012.

[65] K. Park, H. Jee, T. Lee, S. Jung, and H. Lim, “Automatic extraction of user’s search intention from web search logs,” *Multimed. Tools Appl.*, vol. 61, no. 1, pp. 145–162, 2012, doi: 10.1007/s11042-010-0723-8.

[66] S. Yang *et al.*, "Intention mining in medical process: A case study in trauma resuscitation," *Proc. - 2018 IEEE Int. Conf. Healthc. Informatics, ICHI 2018*, pp. 36–43, 2018, doi: 10.1109/ICHI.2018.00012.

[67] H. J., "Cochrane Handbook for Systematic Reviews of Interventions | Cochrane Training." https://training.cochrane.org/handbook/current (accessed Oct. 19, 2022).

[68] CORE Computing Research & Education, "CORE Rankings Portal." https://www.core.edu.au/conference-portal (accessed Oct. 18, 2022).

[69] A. Sureka, "Intention-Oriented Process Model Discovery from Incident Management Event Logs," pp. 1–4, 2015, [Online]. Available: http://arxiv.org/abs/1507.01062

[70] Y. Shang, "Studies on User Intent Analysis and Mining," *ProQuest Dissertations and Theses*. p. 100, 2017. [Online]. Available: https://search.proquest.com/docview/2014424447?accountid=14169

[71] B. Hajer, B. Arwa, H. Lobna, and G. Khaled, "Intention mining data preprocessing based on multi-agents system," *Procedia Comput. Sci.*, vol. 176, pp. 888–897, 2020, doi: 10.1016/j.procs.2020.09.084.

[72] S. S. Date, "A Comprehensive Review on Intents, Intention Mining and Intention Classification," *Int. J. Sci. Res.*, vol. 9, no. 11, pp. 16–20, 2020, doi: 10.21275/SR201015182533.

[73] S. Wu, "An user intention mining model based on fractal time series pattern," *Fractals*, vol. 28, no. 8, pp. 1–12, 2020, doi: 10.1142/S0218348X20400174.

[74] R. Deneckère, C. Hug, G. Khodabandelou, and C. Salinesi, "Intentional process mining: Discovering and modeling the goals behind processes using supervised learning," *Int. J. Inf. Syst. Model. Des.*, vol. 5, no. 4, pp. 22–47, 2014, doi: 10.4018/ijismd.2014100102.

[75] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi, "Process mining versus intention mining," *Lect. Notes Bus. Inf. Process.*, vol. 147 LNBIP, pp. 466–480, 2013, doi: 10.1007/978-3-642-38484-4_33.

[76] G. Khodabandelou, C. Hug, R. Deneckere, and C. Salinesi, “Supervised intentional process models discovery using Hidden Markov models,” *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, 2013, doi: 10.1109/RCIS.2013.6577711.

[77] M. Bajec and M. Krisper, “A methodology and tool support for managing business rules in organisations,” *Inf. Syst.*, vol. 30, no. 6, pp. 423–443, 2005, doi: 10.1016/j.is.2004.05.003.

[78] G. Khodabandelou, C. Hug, R. Deneckère, and C. Salinesi, “Supervised vs. unsupervised learning for intentional process model discovery,” *Lect. Notes Bus. Inf. Process.*, vol. 175 LNBIP, pp. 215–229, 2014, doi: 10.1007/978-3-662-43745-2_15.

[79] A. M. Almasoud, H. S. Al-Khalifa, and A. Al-Salman, “Recent developments in data mining applications and techniques,” *10th Int. Conf. Digit. Inf. Manag. ICDIM 2015*, no. Icdim, pp. 36–42, 2016, doi: 10.1109/ICDIM.2015.7381867.

[80] R. Andrews, S. Suriadi, C. Ouyang, and E. Poppe, “Towards Event Log Querying for Data Quality: Let’s Start with Detecting Log Imperfections,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11229 LNCS, pp. 116–134, 2018, doi: 10.1007/978-3-030-02610-3_7.

[81] K. M. Van Hee, Z. Liu, and N. Sidorova, “Is my event log complete? - A probabilistic approach to process mining,” *Proc. - Int. Conf. Res. Challenges Inf. Sci.*, 2011, doi: 10.1109/RCIS.2011.6006848.

[82] S. Suriadi, R. Andrews, A. H. M. ter Hofstede, and M. T. Wynn, “Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs,” *Inf. Syst.*, vol. 64, pp. 132–150, 2017, doi: 10.1016/j.is.2016.07.011.

[83] M. Jans and P. Soffer, “From relational database to event log: Decisions with quality impact,” *Lect. Notes Bus. Inf. Process.*, vol. 308, pp. 588–599, 2018, doi: 10.1007/978-3-319-74030-0_46.

[84] M. Ross, “The 10 Essential Rules of Dimensional Modeling - Kimball Group.” https://www.kimballgroup.com/2009/05/the-10-essential-rules-of-dimensional-modeling/ (accessed Oct. 18, 2022).

[85] A. B. Rebolj and S. F. Possibilities, “Adrijana Biba Starman The case study as a type of qualitative research,” no. September 2014, 2017.

[86] W. M. Van Der Aalst and S. Dustdar, “Process mining put into context,” *IEEE Internet Comput.*, vol. 16, no. 1, pp. 82–86, 2012, doi: 10.1109/MIC.2012.12.

[87] F. E. Ayo, O. Folorunso, and F. T. Ibharalu, “A probabilistic approach to event log completeness,” *Expert Syst. Appl.*, vol. 80, pp. 263–272, 2017, doi: 10.1016/j.eswa.2017.03.039.

[88] H. J. Cheng and A. Kumar, “Process mining on noisy logs - Can log sanitization help to improve performance?,” *Decis. Support Syst.*, vol. 79, pp. 138–149, 2015, doi: 10.1016/j.dss.2015.08.003.

[89] SIGNAL, “The Signal Media One Million News Articles Dataset.” https://docs.google.com/forms/d/e/1FAIpQLSfsNAD8UNJxi5SLGeBPEgYZdWUQfSDc4r95p8eSjMlsnqibgg/viewform?c=0&w=1 (accessed Oct. 19, 2022).

[90] O. Diaz, “GitHub - OswaldoDiaz975/PythonCode_IntentionMining.” https://github.com/OswaldoDiaz975/PythonCode_IntentionMining (accessed Oct. 18, 2022).

[91] C. D. Manning, J. Bauer, J. Finkel, and S. J. Bethard, “The Stanford CoreNLP Natural Language Processing Toolkit,” *Proc. of52nd Annu. Meet. ofthe Assoc. Comput. Linguist. Syst. Demonstr.* pages 55–60, Balt. Maryl. USA, June 23-24, 2014. c?2014 Assoc. Comput. Linguist., pp. 55–60, 2014.

[92] J. Vasilev and M. Kehayova-Stoycheva, “Sales analysis by the rectangle method,” *Leonardo Electron. J. Pract. Technol.*, no. 30, pp. 149–160, 2017, [Online]. Available: http://lejpt.academicdirect.org

[93] Z. YongMajie, W. Li, and W. Yun, *Fast genetic algorithm for pick-up path*

optimization in the large warehouse system, vol. 9142. 2015. doi: 10.1007/978-3-319-20469-7_5.

[94] S. A. Matrix, "Sales Analysis Matrix" Myob Exo Business White Paper, Myob Enterprise Solutions. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://help.myob.com.au/exo/whitepapers/MYOB%20EXO%20Business%20-%20Subscriptions.pdf.

[95] W. Every and C. Cleaning, "Worth Every Cent Cleaning Sales & Marketing Manual" chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://wortheverycent.net/wp-content/uploads/2019/12/SalesandMarketingManual-RV2-.pdf.

[96] L. AbdilA, "Sales, Marketing, and Customer Service," 2021, [Online]. Available: http://eprints.binadarma.ac.id/5014/1/LeonAbdillah_MA2021_11 Sales%2C Marketing and Customer Service 20210525.pdf

[97] C. Cohan-Sujay and Y. Madhulika, "Intention Analysis for Sales, Marketing and Customer Service.," *COLING (Demos)*, no. December 2012, pp. 33–40, 2012, [Online]. Available: http://acl.eldoc.ub.rug.nl/mirror/C/C12/C12-3005.pdf

[98] M. Martínez, "Validez y confiabilidad en la metodología cualitativa," *Paradígma*, vol. 27, p. 25, 2006.

[99] P. Hayashi, G. Abib, and N. Hoppen, "Validity in qualitative research: A processual approach," *Qual. Rep.*, vol. 24, no. 1, pp. 98–112, 2019, doi: 10.46743/2160-3715/2019.3443.

[100] M. Bryant, "Conducting observational research," *Res. Methods Heal. Promot.*, pp. 107–128, 2006.

[101] National Association of Testing Authorities, "Guidelines for the validation and verification of quantitative and qualitative test methods," *Nata*, no. December 2006, pp. 1–32, 2012, [Online]. Available: http://www.demarcheiso17025.com/document/Guidelines for the validation and

verification of quantitative and qualitative test methods.pdf

[102] S. Bravo-Grau and J. P. Cruz, “Estudios de exactitud diagnóstica: Herramientas para su Interpretación Diagnostic accuracy studies: Tools for its Interpretation,” *Rev. Chil. Radiol. año*, vol. 21, no. 4, pp. 158–164, 2015, [Online]. Available: http://dx.doi.org/10.4067/S0717-93082015000400007.

Appendices

Appendix A. Data dictionary of knowledge base

DATA TABLE	COLUMN	DATA TYPE
articles	doc_text	text
	doc_id	text
sentences	lemmas	ARRAY
	pos_tags	ARRAY
	ner_tags	ARRAY
	sentence_id	character
	tokens	ARRAY
	doc_id	character
	sentence_text	text
strategy_qel	strategy_name	text
strategy_candidate	strategy_name	text
	strategy_id	text
strategy_mention	mention_id	character
	mention_text	text
	doc_id	character
	sentence_id	character
	begin_index	integer
	end_index	integer
strategy_rule	strategy_name	text
	rule_name	ARRAY
	strategy_num	integer
strategy_weight	weight	numeric
	strategy_name	text
user_strategy	strategy_name	text
	ponderation	numeric

Appendix B. Procedure to load articles

```
Open database
Read (aticles200.csv)
For each Article
   Insert (Article to database)
End For
Close database
```

Appendix C. Sentences tokenization and lemmatization procedure

```
Open database
Read (Article)
For each Sentence
          Tokenization (Sentence)
          For each Token
```

```
            Lemmatization (Token)
        End For
        Write (Sentence)
End For
Close database
```

Appendix D. Sentences POS tagging procedure

```
Open database
Read (Sentence)
For each Sentence
        For each Token
            Pos_Tagging (Token)
        End for
End For
Close database
```

Appendix E. Sentences NER tagging procedure

```
Open database
Read (Sentence)
For each Sentence
        For each Token
            Ner_Tagging (Token)
        End for
End For
Close database
```

Appendix F. Strategy mention procedure

```
Open database
For each Sentence
        Sentence_Tokens = [Ner_tag_token]
       Elements_Strategy_Mention = [VERB, NOUN, ADJECTIVE, ADVERB, ADVERB
                    COMPARATIVE, ADJECTIVE COMPARATIVE, ADJECTIVE
                    SUPERLATIVE, ADVERB SUPERLATIVE]
        If len (val For val in Sentence_Tokens If val in Elements_Strategy_Mention) > 0
            Strategy_Mention = Sentence
            Write (Strategy_Mention)
        End If
End For
Close database
```

Appendix G. Candidate strategies extraction procedure

```
Open database
For each Strategy_Mention
        Sentence_Tokens = [Ner_tag_token]
       If (Sentence_Tokens CONTAINS "VERB" or "NOUN VERB" or "ADJECTIVE
          NOUN VERB" or "NOUN VERB ADVERB")
            Candidate_Strategy = TokenSelected (Sentence_Tokens)
            Write (Candidate_Strategy)
       End If
Close database
```

Appendix H. Procedure to load strategies of the QEL (sales business)

```
Open database
Read (QEL.csv)
For each QEL
   Insert (Strategy_QEL)
End For
Close database
```

Appendix I. Procedure to verify the strategies structure

```
Open Database
Read (Candidate_Strategy)
Read (Strategy_QEL)
For each Candidate_Strategy
        Rul_name = []
        If len (Candidate_Strategy == 1) Rul_name.Append (VERB)
        If len (Candidate_Strategy == 2) Rul_name.Append (NOUM VERB)
        If len (Candidate_Strategy == 3) Rul_name.Append (NOUN VERB NOUN /
                ADJECTIVE NOUN VERB / NOUN VERB ADVERB)
        Write (Strategy_rule)
End For
For each Strategy_QEL
        Rul_name = []
        If len (Strategy_QEL == 1) Rul_name.Append (VERB)
        If len (Strategy_QEL == 2) Rul_name.Append (NOUM VERB)
        If len (Strategy_QEL == 3) Rul_name.Append (NOUN VERB NOUN / ADJECTIVE
                NOUN VERB / NOUN VERB ADVERB)
        Write (Strategy_rule)
End For
Close database
```

Appendix J. Procedure to weighting of strategies according to the fulfil the business heuristic rules

```
Open database
Read (Strategy_rule)
Read (Heuristic_rule)
TRADING = ['order', 'quotation', 'stock', 'sale', 'price']
DEALING = ['sell', 'buy', 'offer', 'promotion', 'billing', 'cancel']
CRM = ['customer', 'empathy', 'user', 'ecommerce', 'e-commerce', 'omnichannel', 'omni-
   channel']
For each Strategy_rule
        weight1=0
        If (Strategy_rule == "VERB" or "NOUN") weight1 = 1
        If (Strategy_rule == "NOUN VERB" or "VERB NOUN") weight1 = 2
        If (Strategy_rule == "NOUN VERB NOUN” or “ADEJCTIVE NOUN VERB” or
                “NOUN VERB ADVERB") weight1=3
        weight2=0
        List = len (val for val in “TRADING” if val in Strategy_rule)
        If len (List) > 0
                Strategy_weight.Append ("TRADING")
                weight2 = weight2 + List*3
```

```
        List = len (val for val in "DEALING" if val in Strategy_rule)
        If len (List) > 0
                Strategy_weight.Append ("DEALING")
                weight2 = weight2 + List*3
        List = len (val for val in "CRM" if val in Strategy_rule)
        If len (List) > 0
                Strategy_weight.Append ("CRM")
                weight2 = weight2 + List*3
        weight = (weight1 + weight2) / 12
        Write (Strategy_weight)
End For
Close database
```

Appendix K. Procedure to formalizing the user strategies of sales business

```
Open database
Read (Strategy_weight)
For each Strategy_weight
        If (weight > 0.3)
                Write (User_Strategy)
        End If
End For
```

Appendix L. Procedure to calculus the precision measures

```
Open database
CURRENT_SALES_STRATEGIES = ['billing customer sale', 'cancel customer order',
                    'complete customer order', 'customer service', 'deliver customer order',
                    'emit customer quotation', 'generate customer order', 'local stock control',
                    'home sale delivery', 'register sale', 'remote stock control', 'sales record']
TRADING = ['order', 'quotation', 'stock', 'sale', 'price']
DEALING = ['sell', 'buy', 'offer', 'promotion', 'billing', 'cancel']
CRM = ['customer', 'empathy', 'user', 'ecommerce', 'e-commerce', 'omnichannel', 'omni-
channel'] ## Customer Relationship Management
## Calculus of parameters
TP = 0 ## True Positive. Current sales strategies that also are user strategies
FP = 0 ## False Positive. User strategies that are not current sales strategies
FN = 0 ## False Negative. Current sales strategies that are not user strategies, this will
never happen
TN = 0 ## True Negative. Strategies that are not current sales strategies and either are
not user strategies
for each User_strategies
        if user_strategy in CURRENT_SALES_STRATEGIES
                TP=TP+1
        else
                if user_strategy in TRADING or in DEALING or in CRM
                        FP=FP+1
                else:
                        TN=TN+1
                End if
        End if
End for
print ('TP : ', TP)
print ('FP : ', FP)
```

```
print ('FN : ', FN)
print ('TN : ', TN)
Close database
```