



HAL
open science

Intentional Process Mining: Discovering and Modeling the Goals Behind Processes using Supervised Learning

Rebecca Deneckère, Charlotte Hug, Ghazaleh Khodabandelou, Camille Salinesi

► To cite this version:

Rebecca Deneckère, Charlotte Hug, Ghazaleh Khodabandelou, Camille Salinesi. Intentional Process Mining: Discovering and Modeling the Goals Behind Processes using Supervised Learning. International Journal of Information System Modeling and Design, 2014, Research Challenges in Information Science (RCIS'2013), 5 (4), pp.22 - 47. 10.4018/ijismd.2014100102 . hal-01094781

HAL Id: hal-01094781

<https://paris1.hal.science/hal-01094781>

Submitted on 13 Dec 2014

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

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

Intentional Process Mining: Discovering and Modelling the Goals behind Processes Using Supervised Learning

Rebecca Deneckère, Charlotte Hug, Ghazaleh Khodabandelou, Camille Salinesi
Centre de Recherche en Informatique, Université Paris 1 Panthéon – Sorbonne, France

ABSTRACT

Understanding people's goals is a challenging issue that is met in many different areas such as security, sales, information retrieval, etc. Intention Mining aims at uncovering intentions from observations of actual activities. While most Intention Mining techniques proposed so far focus on mining individual intentions to analyze web engine queries, this paper proposes a generic technique to mine intentions from activity traces. The proposed technique relies on supervised learning and generates intentional models specified with the Map formalism. The originality of the contribution lies in the demonstration that it is actually possible to reverse engineer the underlying intentional plans built by people when in action, and specify them in models e.g. with intentions at different levels, dependencies, links with other concepts, etc. After an introduction on intention mining, the paper presents the Supervised Map Miner Method and reports two controlled experiments that were undertaken to evaluate precision, recall and F-Score. The results are promising since we were able to find the intentions underlying the activities as well as the corresponding map process model with satisfying accuracy, efficiency and performance.

Keywords: intention mining, trace, supervised learning, Hidden Markov model, goal modeling, event log

INTRODUCTION

Process Mining has been a topic of interest that has attracted a growing number of publications for the past 10 years (Tiwari et al., 2008). The need of companies to better know their processes, model them, check their alignment with strategic goals, monitor their evolutions has generated a wealth of applications, from business process monitoring to reverse engineering and software process modeling, which in their turn raise new research issues.

Most of the existing process mining approaches deal with process specified with notations that belong either the activity-driven or to the product-oriented paradigm such as BPMN, EPC, Petri Nets, etc. (van der Aalst, 2011 ; van der Aalst & Weijters, 2004 ; Pérez-Castillo et al., 2011). Although extremely interesting to deal with a number of issues, process models specified with these notations are difficult to exploit when it comes to tracing their rationale and measuring. More importantly, process models specified in this way lack of flexibility (Nurcan, 2008) making difficult their alignment with strategic goals, measuring their degree of variability, or even trying to monitor their underlying strategies. This paper builds upon a thread of research works on intentional process modeling (Yu & Mylopoulos, 1994) (Nurcan

et al., 2005) (Yu, 1995), i.e. where process models are specified notations that belong to the intentional paradigm, in other words goal-oriented process models.

Intentions are a first class concept of Information System (IS) engineering (Rolland & Salinesi 2005). In the early 80s, Intentional models were proposed in the IS community (Swanson, 1982) (Christie, 1981) as a “potential theoretical foundation” to determine users’ behavior (Davis et al., 1989). Intention modeling takes root in a former work (Ajzen & Fishbein, 1975) that introduced the Theory of Reasoned Action (TRA) designed to model human’s behavioral intention. The TRA has proven effective in predicting and explaining human behavior through various domains as consumer behavior... Later on in the early 90s, intention analysis and modeling have been promoted as a driving paradigm to study strategic alignment, to define actors and roles, to specify the outcome of business process models and name them, to guide requirements elicitation, analysis, traceability, to study users behavior to identify and name use cases, etc. If intentions are referred to as goals, then intentional process modeling refers to modeling the goals underlying the studied processes (Kaabi & Souveyet, 2007). Notations used in intentional process modelling, and therefore intentional process mining are thus goal modelling notations (Nurcan et al., 2005) (Yu, 1995).

Several methods were recently proposed to mine intentions from observed behaviors. The key idea is to extract sequences of activities from records to evaluate and predict the users’ intentions that resulted in those activities (Khodabandelou et al., 2013). In these works, intentions are considered as “goals to be achieved by performing processes” (Bonito et al., 2009). As a result, mining intention from process traces or logs can be considered an inverse problem, i.e. drawing intentions backward from process performance.

The mainstream research on intention mining lies in the domain of information retrieval (Jathava et al., 2011), (Baeza Yates et al., 2006) (González-Caro & Baeza-Yates, 2011), (Hashemi et al., 2008), (Sadikov et al., 2010), (Strohmaier & Kröll, 2012), (Zheng et al., 2002). Other applications have also been published, e.g. contents analysis (Mei et al., 2005), or business process models improvement (Outmazgin & Soffer, 2013). The common characteristic of the aforementioned methods is that they almost systematically generate individual intentions. This is interesting, but in the context of IS Engineering, intentions must be modeled, and dependencies between them and other concepts such as resources, tasks, strategy, systems functions, etc. be specified. The main contribution of this paper is a mining method that produces intentional process models, i.e. conceptual models of the intentions behind processes.

We believe new applications will be found in the near future. Intentional process mining might help improving guidance, provide better recommendations, facilitate process modeling and process model quality assessment, identify the gap between prescribed business requirements and goals, help CEOs assess and monitor strategic goal implementation, etc.

In the context of Information Systems (IS) engineering, intentional process mining can be useful at different stages of the process model lifecycle, for instance (i) at the requirements level, to elicit actual users' goals rather than inferred ones, (ii) at the project management level, to check the alignment between a prescribed objectives and the actual processes model or (iii) at the application level, to supervise users activities and provide them with more useful recommendations at runtime.

This paper presents the Supervised Map Miner Method, an intentional process mining method that uses supervised learning and Hidden Markov Models to generate intentional models specified with the Map formalism (Rolland et al., 1999). Besides relating process intentions to each other, Map process models address the optative nature of intentions (Jackson, 1995) by specifying the variable strategies that are used to achieve them through the variety of processes enactments. The main contribution of this paper is that the mined intention process models make explicit and formalize the intentions that underlie the mined process traces. Rather than just indicating the category of intentions, or worse, simply state that there is an intention, the Supervised Map Miner Method labels the mined intentions and strategies with a human-understandable name and organizes them in conceptual models specified with a formal syntax. With respect to our former works on intentional process mining, this paper presents two controlled experiments performed with two different intentional process models.

The rest of the paper is structured as follows: section 2 introduces the domain of intention mining (to which our own approach compares better than more traditional process mining techniques), and provides an overview on the Supervised Map Miner Method. Section 3 provides further details on the Supervised Map Miner Method. Section 4 reports the two experiments. Impact on theory and practices are discussed in section 5.

THEORETICAL BACKGROUND AND RELATED WORKS

A quick search in the literature reveals that (a) many intention mining techniques have already been proposed, and (b) this research area is extremely dynamic with new contributions continuously published. Rather than aiming at a systematic literature review, this section first introduces the area by describing three particular approaches that were selected because of their impact or originality: (Strohmaier & Kröll, 2012), (Baeza et al., 2006) and (Outmazgin & Soffer, 2013).

Selected Works on Intention Mining

Strohmaier & Kröll, 2012

Strohmaier and Kröll's method is one of the many approaches to acquire knowledge about human intentions (the word used here is "goal") by investigating web engine query logs. The idea is that better understanding the rationale behind the actions of web engine users can be useful to deal with a range of issues such as recognizing users' intentions, reasoning about them, or generating plans to help them achieve their intentions. A salient feature of this approach is that it differentiates between implicit and explicit intentions.

Explicit intentions are specified or expressed by people. They can for instance be expressed in natural language, in a sentence predicate with an intention verb and complements, as (Prat, 1997) describes it. In query engines, explicit intentions can appear in queries such as "I want to do X", X being the intention that the web engine user wants to achieve. Even better, the query is sometimes simply the intention itself: "X" such as in "plan a menu for diner".

Implicit intentions underlie what is expressed by people or can be observed from them. Contrary to explicit intentions, implicit intentions are neither expressed or specified as such, nor indicated e.g. by selecting among a list of intentions that a system helps to achieve. They must therefore be labeled to become explicit, which is the purpose of Strohmaier and Kröll's approach.

The approach has 3 main stages. First, word-unigrams and part-of-speech analysis algorithms are used to detect the main intentional features of web search queries. Then, a classification is achieved, which can be done using different algorithms such as SVM or Naïve Bayes Classifier. Last, Levin's verb classes (Levin, 1993) are used to identify the class of intention to which the web search queries correspond. The WEKA (Witten & Frank, 2005) and the Natural Language Toolkit (NLTK) tools are used to manipulate the web search queries and to determine intention classes.

Baeza-Yates, Calderón-Benavides & González-Caro, 2006

The purpose of Baeza-Yates' et al., approach is to discover implicit intentions (called "query intents") from web engine queries. According to Baeza-Yates' et al, mining query intents is important for web engines; the idea is that query results can be of better quality when the intentions behind queries are better understood. Baeza-Yates' et al., approach relies on a fundamental distinction between informational or non-informational intentions.

Informational intentions aim at acquiring some kind of knowledge, information, data, etc. The purpose of non- informational intentions is to perform some kind of process, task, transaction, action, activity, etc. Of course informational and non-informational intentions are not disconnected: achieving some non-informational intention (such as cooking) may require to search for information (a recipe), and the other way round, achieving a higher informational intention (getting the list of ingredients) may involve some lower level non-informational intention (converting weights from the imperial to the metric system). Things are not always clear-cut. An intention is called ambiguous when it is neither clearly informational nor non- informational.

In Baeza-Yates' et al., approach, query intents are mined from event logs of individual web engine users. The proposed method mixes supervised and unsupervised machine learning using a classification algorithm, together with predefined intention categories from the Open Directory Project. The method uses the SVM algorithm combined with Error-Correcting Output Coding (ECOC), and Probabilistic Latent Semantic Analysis (PLSA) to analyze users' interests and classify them into categories. Its implementation relies on the PennAspect software (Schein et al., 2001) to retrieve intentions names.

Outmazgin & Soffer, 2013

First of all, Outmazgin and Soffer's approach does not explicitly deal with intention but on the concept of "workarounds". *Workarounds* are non-compliant behaviors. They can be observed when people deviate -in full knowledge- from the processes they are supposed to follow. The driving idea is that detecting and understanding workarounds make it possible to improve process models.

Outmazgin and Soffer's method exploits workaround patterns that are first created through qualitative studies performed in different organizations. Actual workarounds are detected using the Disco tool (van der Aalst, 2011) that generates models specified using the Business Process Model Notation (OMG, 2011).

Besides the original use of process mining techniques, what is interesting here is to notice that the concept of workaround draws a clear link between process models and intentions. First, workaround patterns point out to implicit intentions. Second, the method relates these "actual" implicit intentions (from peoples' behavior observations) with "theoretical" explicit intentions (from process models). In brief, the intentions mined in this method stay implicit: they are spotted but never named. Nevertheless, it is indeed on their analysis that the technique relies.

SUPERVISED MAP MINER METHOD

The Supervised Map Miner Method was designed to mine intentional models from traces of non-deterministic activities that follow a stochastic process. Two important assumptions are that (i) intentions are at multiple levels, higher-level intentions embed lower level ones, and (ii) the multi-level structure of users' activities and intentions is a time-variant, i.e. the system output evolves over time and non-linear system, i.e., a system that has more than one dimension and the output is not directly proportional to the input. HMMs are particularly adapted to handle this kind of situation.

A HMM can be considered as simpler Dynamic Bayesian Network (DBN) (Murphy, 2002), i.e. a Bayesian Network (BN) that models time series data and relates variables to each other over adjacent time steps. They are also called two-time slice BN because at any point in time t , the value of a variable can be calculated from the immediate prior value (time $t-1$). It is important to notice that DBN makes the fundamental assumption that events can cause other events in the future but not in the past. At first glance, this assumption fits rather well with intentions but it can be argued. However, whereas future intentions influence past ones is a philosophical concern outside the scope of this paper.

Overview of the Supervised Map Miner Method

Before introducing the Supervised Map Miner Method, it is important to precise its position regarding the literature, more particularly regarding the three approaches explained in section 2.1. We compare the characteristics of the aforementioned approaches with respect to four essential aspects: the type of inputs, the analysis paradigm, the methods used and the type of outputs.

- Whereas the inputs of these approaches are single queries, logs, or activities, the input of our approach is the temporal set of a user's activities – the interactions of a user with a computer system during a time slice $\Delta = t_N - t_0$, where t_0 is the beginning of the activity performance and t_N is the end. During a time Δ one or several sequences of activities are recorded by a tool. These sequences are a trace of activities for the user.

- Since the inputs the mentioned approaches are individual entities, their analysis does not take into account the temporal dependencies between each input element. In our approach, we analyze the entire users' traces that occurred during a process enactment. Note that a process in the IS context is defined as a sequence of activities linked to each other by a common goal (Rozinat, 2010). Therefore, there are strong correlation and dependency between users' activities and they cannot be considered as a single, independent and uncorrelated entity. A sequence of activities contains richer information about the users' intention than a single activity, and this from both semantic and abstraction level points of view. Indeed, analyzing a sequence of activities allows determining the high-level intentions (e.g. organizational goal), while analyzing single activities leads to less informative, low-level intentions, also called basic intentions or action intentions, which are closer to activities than intentions.

- These approaches use classification techniques to classify a single input into a class of intention. The choice of classification technique seems accurate since, as explained above, their concept of intention is different from our concept. The aim of our approach is to determine users' strategies and intentions and formalize them by an adequate intentional process model. Thus, we set up a two-level topology which discovers the strategies first, and then the intentions. Note that this hierarchy corresponds to the human reasoning. Thus, we need to model the inputs with a mathematical model that supports this topology.

- The outputs of these approaches are the low-level intentions, which mean inferred intentions are not part of the high-level intentions that a user wants to fulfill in a given process. Whereas in our approach the outputs are the high-level intentions of a process that can be discovered from a set of sequences of activities linked to each other by a common objective.

Besides identifying intentions, the object of intention mining approaches may be to generate intentional models specified using a formal notation. Focusing on the IS engineering literature, we distinguish KAOS (Dardenne et al., 1993), i* (Yu, 1995) and Map.

KAOS supports variability and has a well-structured semantic but it has a rigid task-decomposition; modeling complex intentional processes is then difficult. i* has an operational semantic for the tasks but not for the goals and it is not used to model strategic goals. i* only allows modeling the business strategy with organizational strategic goals. i* is not designed to be a variable framework, therefore, it does not afford a high level of flexibility. Tropos (Bresciani et al., 2004) is an agent-oriented software engineering methodology that includes i*; GRL (Amyot et al., 2009) is an intentional modeling language based on a subset of i*. Thus, they both have the same limits as i*.

We chose Map notation rather than other intentional process models notations because (a) its topology, i.e. combining intentions and strategies at different abstraction levels, allows to handle large-scale and complex processes (Rolland and Salinesi, 2005), (b) it supports process variability and flexibility by defining different strategies to fulfill a given intention (Rolland et al, 1999), and (c) it has proved to be effective to specify business processes, engineering methods, software engineering processes, etc. (Rolland et al, 1999) (Rolland, 1993) (Rolland and Salinesi, 2005). A systematic review of these models is discussed in a previous paper (Khodabandelou et al., 2013).

The Map formalism (Rolland et al, 1999) combines the concepts of “intention” and “strategy” in collections of models (called “Maps”) organized hierarchically with refinement links. Intentions are considered as goals to be achieved by performing a process. Strategies define different ways to achieve a given intention. A map (an instance of Map metamodel) specifies the multiple ways of enacting a process to fulfill a given intention with a given strategy. A map is presented as a directed graph where nodes are “intentions” and edges are “strategies” (see Figure 3 section 4.1). A strategy connecting to intentions is called a “section”, which is formally defined by the triplet $\langle \text{Source Intention, Target Intention, Strategy} \rangle$. The sections of a map can be executed as long as intentions are not completely fulfilled. In order to realize the strategies, one has to perform some activities which are recorded as the users’ traces. Specifying different strategies for every intention emphasizes the optative nature of intentions: there are often many ways to achieve intentions.

To the best of our knowledge, two mathematical models are used in intention mining methods: the Bayesian model (Strohmaier & Kröll, 2012) and the Hidden Markov model (Sadikov et al., 2010). *Bayesian networks* are graphical models that represent random variables and their conditional probabilities via a directed acyclic graph. *Hidden Markov Models* (HMM) are a variant of stochastic Markov chain that represent hidden sequences of states, which are interpreted as intentions in the context of intention mining. HMM generalize finite-state automata by evaluating both the probability of transitions between states and probability distributions of observations in those states. The Supervised Map Miner Method uses HMM to represent the probabilities of sequencing intentions and strategies while enacting a process, that is, to model the relation between sequences of strategies and sequences of activities.

Supervised learning is a machine learning technique that consists in inferring a function from labeled training data (Mohri et al., 2012). In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the

supervisory signal). *Unsupervised learning* operates on unlabeled data - input where the desired output is unknown. In the context of intention mining, this can be achieved through cluster analysis. The problem of unsupervised learning is that since the examples given to the learner are unlabeled and there is no error or reward signal, the intention names inferred are hardly validated. For the Learning phase, the Supervised Map Miner Method uses the Viterbi Algorithm (VA) (Forney, 1973).

Figure 1 provides an overview of the Supervised Map Miner Method which is a method of intention mining. As the figure shows, the method consists of several stages: (i) using recorded traces of activities obtained from practical uses of the prescribed map labeled by related strategies as training data, (ii) estimating the parameters of the HMM, i.e. emission and transition matrices, based on the recorded traces, (iii) predicting strategies and intentions for a new trace of activities using VA; last (vi) based on estimated strategies and intentions and on the emission and transition matrices, the Supervised Map Miner Method generates intentional model formalized with the Map metamodel (on the right hand side of the figure). The first two stages are part of the learning phase and the two last stages composed the discovery phase of the method.

The obtained map can be compared with the prescribed map to check the conformance between the models. This can be helpful for: (a) verifying if stakeholders followed the prescribed map or not (b) when and why they deviated from the prescribed map (c) analyzing stakeholders' behaviors during process enactment, (d) enhancing the prescribed map model regarding the results of stakeholders' behavioral analysis, (e) guiding stakeholders in each step of process enactment at runtime (this point is one of the perspectives of our work).

The intentions specified as output in map process models are fully explicit. They aim at providing a view on intentions. Indeed, research on guidance in method engineering shows that many method engineering issues, such as rigidity or lack of adaptation, are solved more effectively when intentions and strategies are explicitly specified (Rolland et al., 2005).

The Supervised Map Miner Method was initially developed for conformance analysis in the method engineering domain (as shown later in this paper). Note that intention mining can be applied for one user or many users to represent the point of view of one user or a group of users.

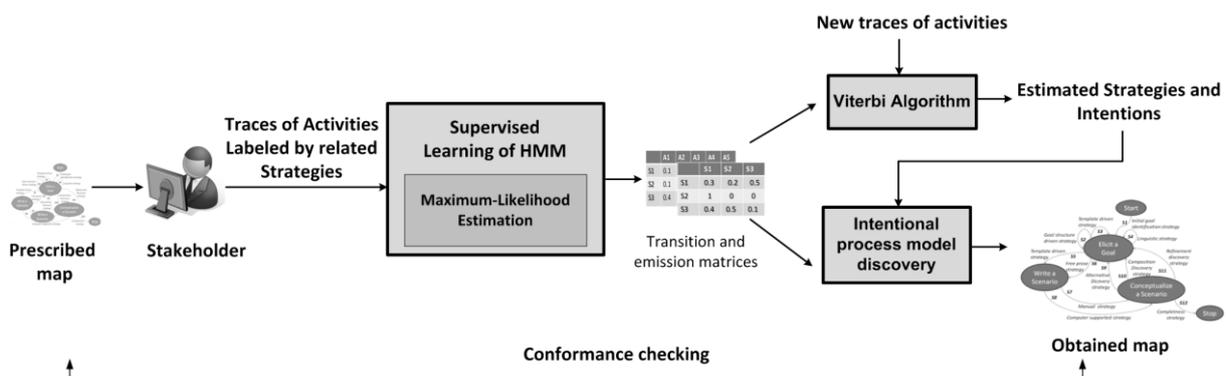


Figure 1. The overview of the Supervised Map Miner Method.

The Supervised Map Miner Method is supervised and semi-automatic. Its inputs are traces, as opposed to event logs used in the aforementioned approaches. A trace consists of several events such that each event relates to the enactment of an activity. Performing one or several activities relates to the enactment of a strategy and consequently the fulfillment of an intention. Last, although the Supervised Map Miner Method automatically discovers the topology of the map process models, i.e. the strategies, it does not still exploit taxonomy or ontology.

Adapting HMM to intention mining

Here we give a brief introduction to HMMs framework. A complete overview of HMMs can be found in (Rabiner, 1989). HMMs consist of two correlated Markov processes. One process is for the hidden state of the HMM and the second process is for the observation. An HMM is formally defined by a tuple: $H = \{S, A, T, E, \pi\}$ where S is the set of possible hidden states, A is the set of possible observations, T is the hidden states transition matrix, (i.e. the matrix that represents the probabilities of transition from one hidden state to another), E is the observations emission matrix (i.e. the matrix which contains the probabilities of observations for every hidden states), and π is the vector of initial probabilities of the hidden states (i.e. the hidden states probabilities at the beginning of the process).

Figure 2 depicts an example of HMM that contains three possible hidden states $\{S_1, S_2, S_3\}$ and three possible observations $\{A_1, A_2, A_3\}$. For instance $T(S_3, S_2)$ represents the transition probability from hidden state S_3 to S_2 and $E_3(A_2)$ represents the emission probability of observation A_2 in hidden state S_3 . From an initial hidden state given by π , an observation is generated according to E , then and for each step of the process a new hidden state is generated according to T and a new observation is generated according to E .

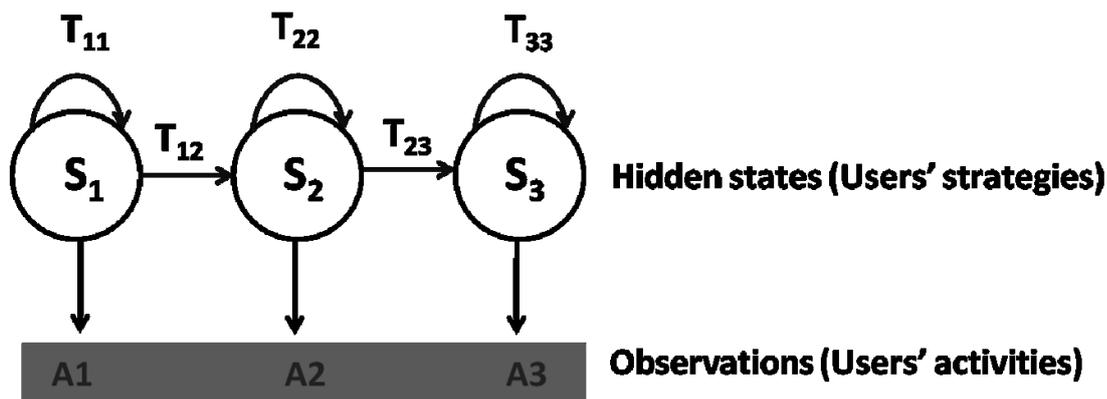


Figure 2. Example of a first order HMM structure with three hidden states.

In our previous work (Khodabandelou, 2013), we proposed to model the hidden states of HMMs as the users' intentions and the observed process as the users' activities. Nevertheless, we realized that this model was not accurate enough to represent precisely a map process model, i.e. instance of the Map metamodel. Indeed, activities are generated by strategies and not directly by intentions. For this reason we propose, in this framework, to model the hidden states as strategies and the observations as users' activities. Note that this model allows inferring intentions since when a sequence of strategies is known; related intentions can also be inferred.

Using HMMs raises several questions: (a) how to estimate the parameters of the HMMs? (b) What is the probability of a given sequence of activities? (c) What are the most probable strategies associated to a given sequence of activities? The first and the third questions are addressed in the following section.

Map mining process

Once the two Markov processes are defined, the parameters of the HMM must be estimated. This is the learning phase of the Supervised Map Miner Method. It consists in training the algorithm with sequences of activities and related sequences of strategies to find the

parameters of the HMM. A second phase is the discovery phase, in which strategies associated to any sequence of activities can be evaluated.

Learning phase: the estimation of the HMM parameters consists in finding the transition probability between strategies (matrix T) and the probability distribution of activities for each strategy (matrix E). If a sequence of activities $A_{1:L}$ is in the model and the corresponding strategies are known, the probability estimates are computed using the Maximum-Likelihood Estimation (Gales, 1998). This method estimates the parameters $T(u, v)$ and $E_u(a)$ such that they analytically maximize the likelihood of co-occurrence of the strategy $S_{1:l}$ and of the sequence of activities $A_{1:l}$. It consists in counting the number of transitions from one strategy to another and the number of occurrences of each activity in each strategy. This learning phase is necessary to evaluate the most likely strategy associated to a given activities sequence.

Discovery phase: Once the parameters of the HMM are estimated, we have to identify the strategy the most associated to a given sequence of activities. To do so, the VA is commonly used in the context of HMM; this algorithm radically reduces the complexity of the search for the most likely hidden sequence of strategies. Thereby, the exponential complexity of a brute force search becomes linear.

EVALUATION

In order to evaluate the precision and recall of the Supervised Map Miner Method, we conducted two controlled experiments with Master students majoring in computer science.

First Controlled Experiment: the E/R case

In the first experiment, a map specifying intentions and strategies for Entity/Relationship modeling was given to the students as guidance. In order to get traces, we developed a web-based tool that records which sections of the map were followed by the students while creating an Entity-Relationship diagram. The traces from 66 students were collected during the experiment. Table 1 presents the profile of the students. All of them had learned and practiced ER modeling since more than one year before the experiment.

Table 1. Profile of the students

Total	Average age	Sex		Master degree	
		Male	Female	1st year	2nd year
66	24,4	49	17	48	18

The intentional model used to guide the students was adapted from (Assar, 2000). The model, presented in Figure 3 shows that students were allowed to act with three intentions in mind: *Specify an entity*, *Specify an association* and *Stop*. According to this model, students can select ten strategies to fulfill the three intentions. Each edge represents a strategy that a student can select to fulfill an intention (specified as a node) according to his/her situation. For instance, if the current situation is *Start* and the students' intention is to *Specify an entity*, there is only one strategy (*by completeness of the model*) to fulfill this intention. When the current situation is *Specify an entity*, there are four strategies (*by completeness, by generalization, by specialization, by normalization*) to fulfill the same intention. It is possible to continue enacting the process by selecting the strategies that lead to the considered intentions but once the *Stop* intention is achieved, the enactment of the process is finished.

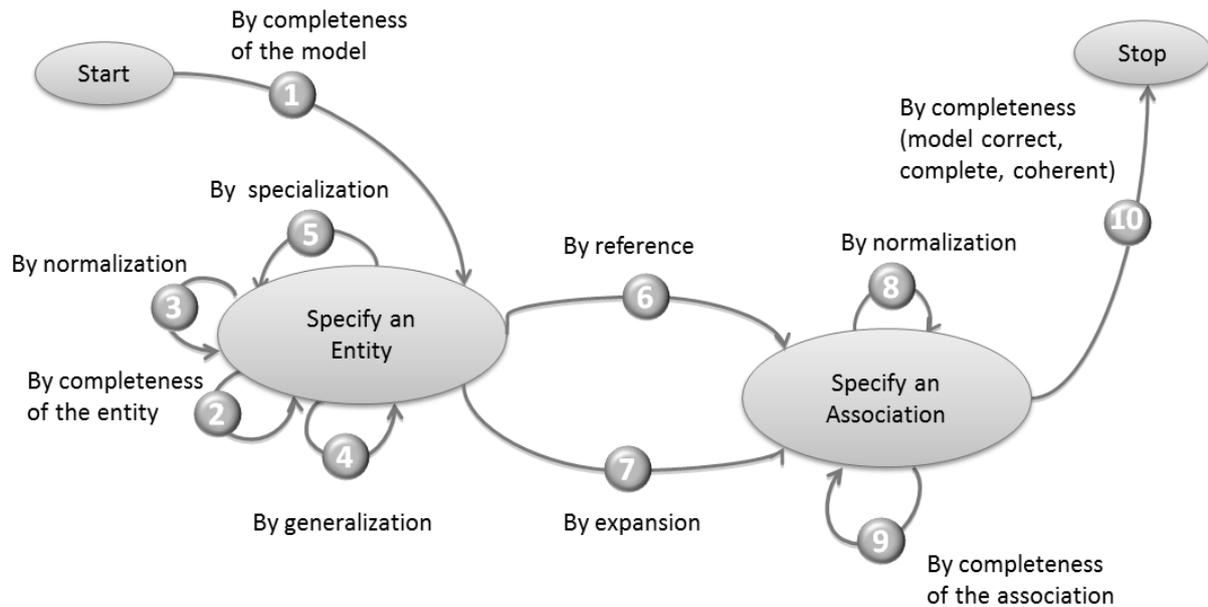


Figure 3. The prescribed intentional model.

To fulfil the recommended intentions following the proposed strategy, students could carry out fifteen different modelling activities named in Table 2. The links between each section of the Map and the relating activities are detailed in Table 3.

Table 2. Observed activities.

Activity	Code	Activity	Code	Activity	Code
Create entity	A1	Link association to entity	A6	Delete entity	A11
Create association	A2	Create generalization link	A7	Delete association	A12
Create attribute	A3	Create specialization link	A8	Check consistency	A13
Link attribute to entity	A4	Define primary key	A9	Check completeness	A14
Link attribute to association	A5	Delete attribute	A10	Delete Link	A15

Table 3. Strategies and related activities.

Strategy number	Strategy	Related activities	Activities codes
1	By completeness (model)	Create entity	A1
2	By completeness (entity)	Create attribute Link attribute to entity	A3, A4
3	By normalization	Delete attribute Delete Link attribute to entity Delete entity Delete attribute * (Delete association, Delete attribute *) * Define primary key	A10, A15 A11, A10*, (A12, A10*)*
4	By generalization	Create entity Create generalization link	A1, A7
5	By specialization	Create entity Create specialization link	A1, A8
6	By reference	Delete attribute Create entity Create association Link association to entity Link association to entity	A10, A1, A2, A6, A6
7	By expansion	Create association	A2
8	By normalization	Delete association (Delete attribute, Delete Link attribute to association)* Delete attribute	A12, (A10, A15)* A10
9	By completeness (association)	Create attribute Link attribute to association	A3, A5
10	By completeness (final)	Check consistency Check completeness	A13, A14

Note in Table 3 that several activities relate to several strategies. This is for instance the case of ‘delete attribute’ activity that relate to both intentions ‘Specify an entity’ and ‘Specify an association’ when they are performed ‘by normalization’ (strategies S_3 and S_8). Indeed, a given activity can be performed in the context of different strategies to achieve different intentions.

The web-based tool traced information about the activities that were executed, the strategies that were selected, a timestamp, and the unique identifier of the students. The model used to store the traces is presented in (Hug et al., 2012).

Strategies discovery

The traces were thus used to estimate the HMM parameters. As explained earlier, this consists in estimating the coefficients of the matrices A and S . Since the prescribed intentional model comprises 10 strategies and 15 activities, the size of S is 10×10 and the size of A is 10×15 . The coefficients of the transition matrix S were obtained by counting the number of transitions from one strategy to another and the coefficients of matrix A , by counting the number of times each activity appeared for each strategy.

The quality of the estimated coefficients depends on the length of the sequences used to calculate the estimates. If the length of the sequences is too short, the sequences will not capture all the typical students' behaviors and the estimated coefficients will be of poor quality.

This phenomenon can be observed on the dataset of the experiment. First, the matrices \hat{E} and \hat{T} were estimated with the full length of the sequence. Then, for seven different lengths of sequences (10, 20, 30, 40, 50, 60 and 66), matrices \hat{E} and \hat{T} were again estimated. Six pairs of matrices of different quality were thus obtained. The quality of estimations is shown in Figure 4. For each sequence length, the figure reports the mean of the absolute values between the coefficients differences for the estimated matrices \hat{E} and \hat{T} with regard to the coefficients of the matrices estimated with the full sequences of length 66 – let's call them \hat{E} -Best and \hat{T} -Best. More precisely, this is the mean of the absolute values of the coefficients differences of these matrices. The figure shows that the coefficients converge as the number of traces for training increases. In other words, the error of estimation of coefficients decreases with the length of the training sequences. The use of the \hat{E} -Best and \hat{T} -Best sequences should therefore be used as the emission and transition matrices.

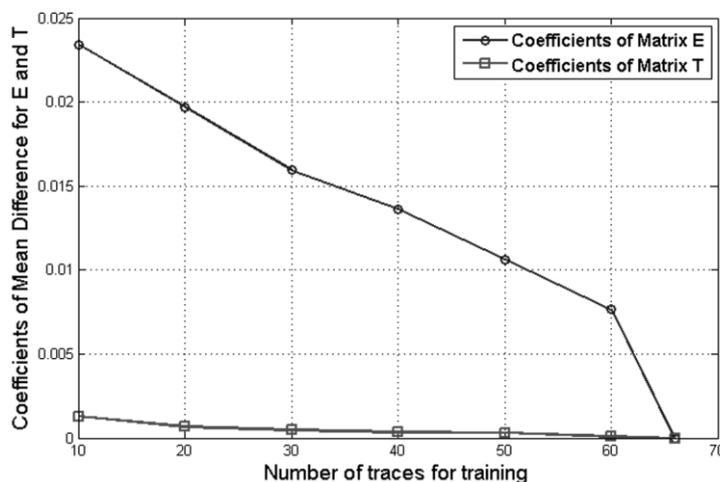


Figure 4. Error of parameters estimation depending on sequence length.

Intentional model discovery

Figure 5 illustrates the intentional model obtained using the best estimated parameter of HMM (the transition matrix) obtained previously. This model shows all the transitions between the intentions in the Map through the sub-intentions defined by SI-1 to SI-8. This intentional model is obtained using the threshold adjusted to $\epsilon=0.06$. This threshold allows adjusting the complexity of a map process model. When ϵ is close to 0, almost all the transitions from the supervised model are present in the obtained map. Consequently, the likelihood of the obtained map is high but the obtained map is more complex since it has a lot of sections. However when ϵ increases the number of sections, as well as the likelihood of the obtained map, decrease. The map in this case is less complex but it is not accurate enough. We conclude from this study that the value of ϵ has to be set to obtain a compromise between the accuracy of the map and its understanding complexity.

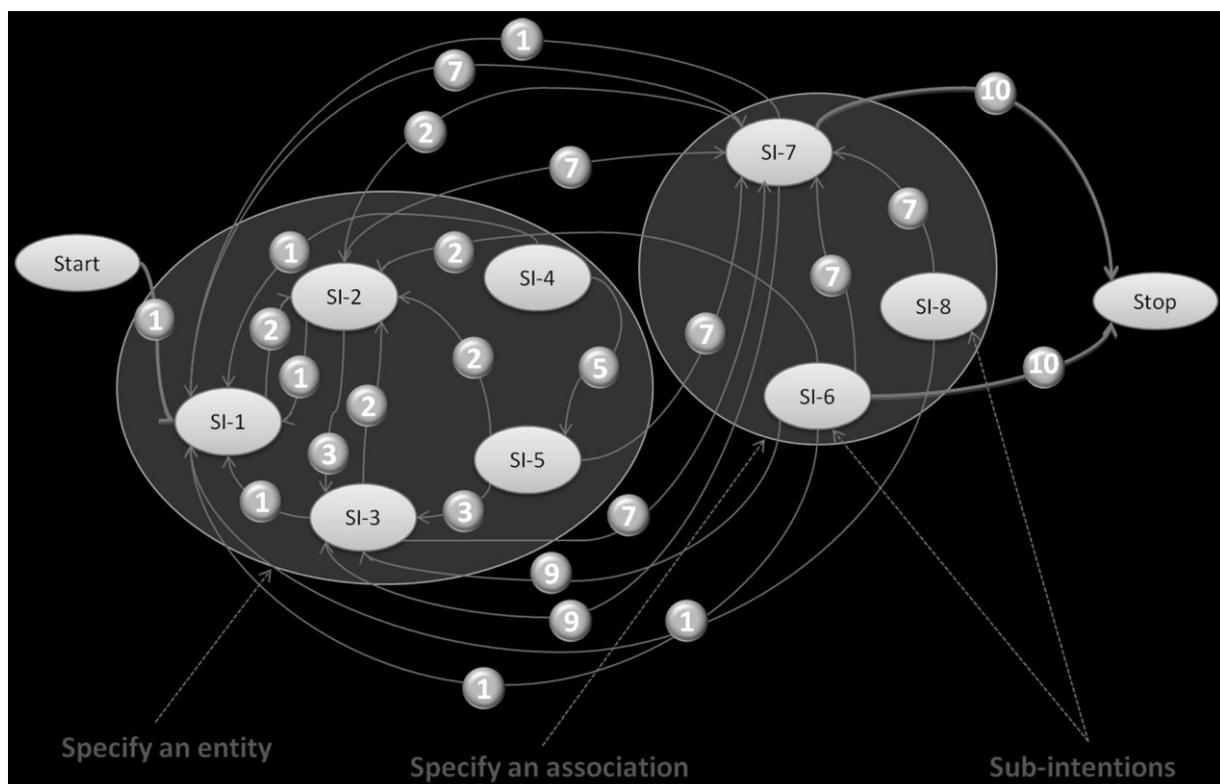


Figure 5. Map obtained using HMM based supervised learning with a threshold $\epsilon=0.06$.

This complex map can be simplified by identifying the prescribed intentions, i.e. *Specify an entity* and *Specify an association* through regrouping manually the sub-intentions into higher level ones. Since our assumption is that the students have respected the prescribed map, then we know to which intentions the strategies found in the obtained map correspond.

For instance, regarding the prescribed map, the strategies 1, 2, 3 and 5 should be associated with the *Specify an entity* intention. Strategy 7 should be defined between these two intentions. Figure 6 shows the final map thereby obtained.

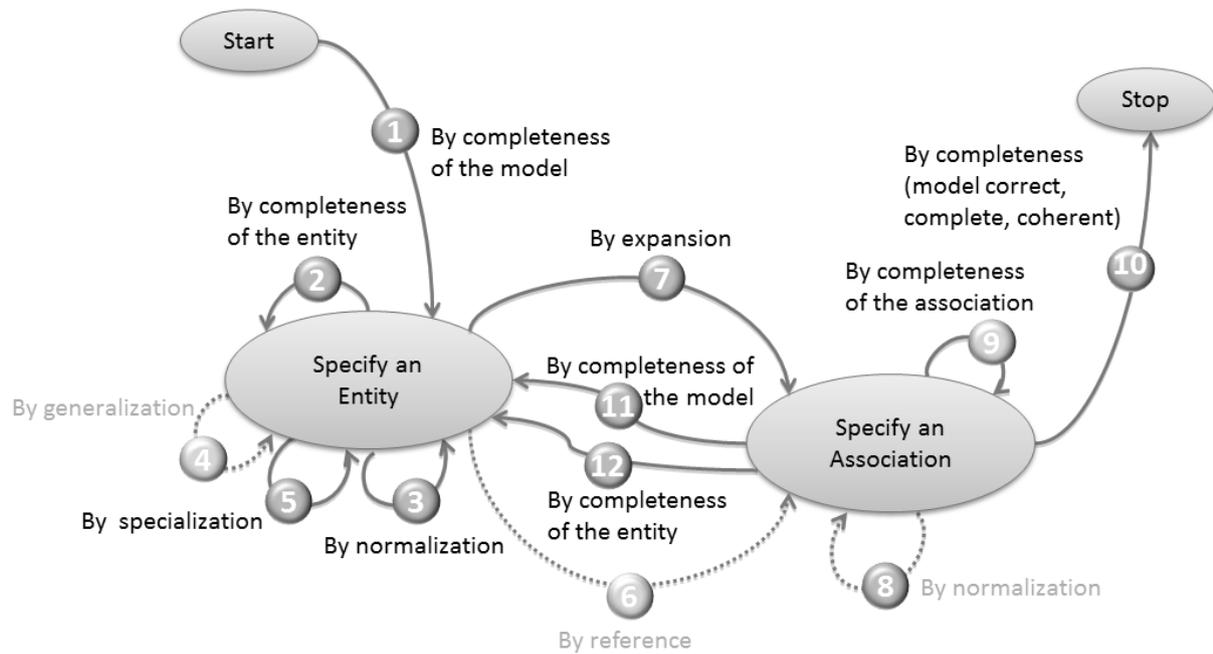


Figure 6. High-level abstraction of the obtained map.

Figure 6 highlights the matches and mismatches between the obtained map and the prescribed one. It means that sometimes the students achieved their intentions respecting the prescribed map and sometimes they deviated from it.

- The first ascertainment concerns the matches between the prescribed and obtained maps. The students chose strategy S_1 to achieve the intention *Specify an entity* and they continued to try to reach this intention by choosing strategies S_2 , S_3 and S_5 . To fulfill the intention *Specify an association*, they chose strategy S_7 and finally strategy S_{10} to *Stop* the process.

- The second ascertainment concerns the mismatches between the prescribed and obtained maps. The students never enacted some prescribed strategies. Dashed arrows in Figure 6 show these strategies. For instance, we observe that they never chose strategies S_4 : *By generalization*, S_6 : *By reference*, S_8 : *By normalization*.

- Some strategies shown in Figure 5 raise some issues. The section <Specify an association, Specify an entity, By completeness of the association> (S_9 in Figure 5) is not coherent as the target intention completely differs from what is implied by the strategy. We then chose not to represent it in Figure 6. The section <Specify an association, Specify an entity, By completeness of the model> was kept in Figure 6 (S_{11}) as users can complete the model by creating a new entity after specifying an association. In this case, the situation takes into account the modified product (the ER diagram), on the contrary of the section <Start, Specify an entity, By completeness of the model> which implies no existing product to take into account. However, we are not able to tell from the traces if the users properly followed these strategies according to the product situation (existing ER diagram or not). With the same reasoning, we kept the section <Specify an association, Specify an entity, By completeness of the entity> in Figure 6 (S_{12}) allowing users to add attributes to an entity of their diagram after specifying an association.

Results Analysis

We calculate the recall, precision and F-score of the Supervised Map Miner method. Five strategies had a 100% score which mean they could systematically be correctly retrieved from

observations of students' activities. The discussion below focuses on the remaining 5 strategies.

The first observation is that recall and precision are stabilized when the estimation sequence length reaches its maximum value. It means that for a length of 66, the VA provides stable results.

Table 4 shows recall, precision and F-score for 5 strategies mined from the 66 traces. For example, the algorithm finds 99% of activities related to strategy S_1 . This means that almost all the activities associated to strategy S_1 were identified. Now the question is: *does the algorithm associate several activities to the strategies while, in fact, they belong to other strategies?* This question could be addressed using the precision ratio. For example, the precision result stabilizes at 99% for strategy S_1 , which means only 1% of the activities are associated to strategy S_1 while they should not.

Table 4. Recall, precision and F-score for 5 strategies.

Strategy	Recall	Precision	F-score
1	99%	99%	0.99
3	95%	92%	0.94
4	72%	65%	0.68
5	87%	92%	0.90
8	73%	78%	0.75

The same measures were taken for all the strategies and reported in Figure 7, which shows the global performance of traces estimation. In this experiment, the curve of recall starts from 0.9114 and its value increases until reaching the maximum value at 0.9240. The curves of precision and F-score start from 0.9120 and 0.8869 and stop at 0.9194 and 0.9207, respectively. We deduce from these results that the accuracy of retrieval for the ten strategies is 0.9207, which indicates that the method has found the right strategies corresponding to the traces of activities with a reliability of 92%.

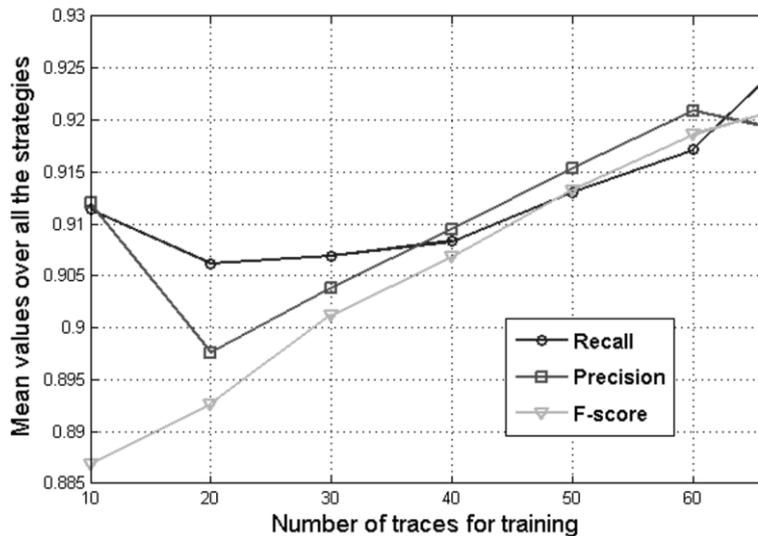


Figure 7. Mean values for recall, precision and F-score for the obtained map.

Second Controlled Experiment: The gift case

The application fields of intentional models are extremely variable. For this case study, we guided students through a prescribed intentional model to buy a present for their best friend. Figure 8 presents this model.

One the reason motivating to conduct the second controlled experiment is to better understand why the students deviated from the first prescribed map. To verify this phenomenon, the second prescribed map offers more strategies to fulfill the intentions than the first one. This makes the prescribed map more flexible and permits the students to pursue their objective more easily.

Context

The prescribed map is composed of three intentions *Find an idea for a present*, *Find a place to buy the present* and *Stop*, 15 strategies, 29 sections and 22 activities (see Table 5).

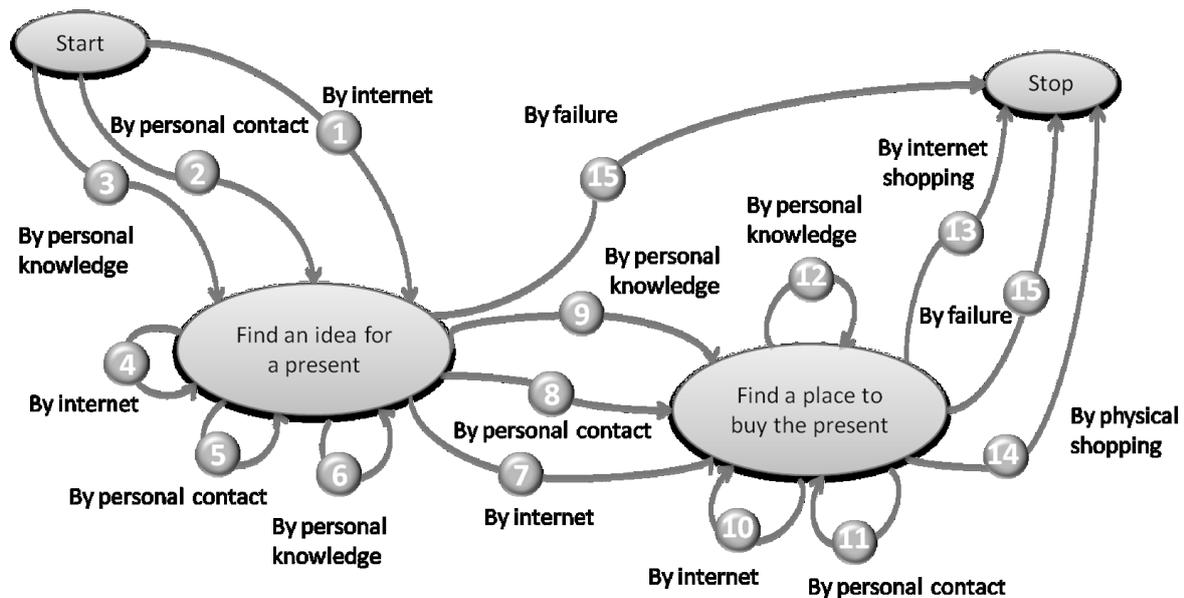


Figure 8. The prescribed intentional model to guide users to buy a present.

Table 5. The observed activities.

Activity	Code	Activity	Code
Open browser	A1	End call	A12
Type query	A2	Choose message menu	A13
Read result	A3	Type SMS	A14
Create idea concept	A4	Send SMS	A15
Type address	A5	Modify idea concept	A16
Navigate on the web site	A6	Create place concept	A17
Open phone	A7	Modify place concept	A18
Open contacts	A8	Online payment	A19
Choose contact	A9	Close browser	A20
Call	A10	Close browser	A21
Speak	A11	Give up	A22

Table 6 presents the strategies and the related activities. Several strategies are grouped in a guideline to ease the reading and understanding of the prescribed map. For instance, the strategy 'by internet' can be performed 'by Google search', 'by website search' or 'by visiting a forum'.

Table 6. Strategies and related activities.

Guideline number	Guideline	Strategy	Corresponding activities
1	By internet	By Google search	A1, A2, A3, A4
		By website search	A1, A5, A3, A6*, A4
		By visiting a web forum	A1, A5, A6*, A4
2	By personal contact	By asking the concerned person	A4
		By calling a friend	A7, A8, A9, A10, A11, A12, A4
		By sending an SMS	A7, A13, A9, A14, A15, A4
3		By personal knowledge	A4
4	By internet	By Google search	A1, A2, A3, A16
		By website search	A1, A5, A3, A6*, A16
		By visiting a web forum	A1, A5, A6*, A16
5	By personal contact	By asking the concerned person	A16
		By calling a friend	A7, A8, A9, A10, A11, A12, A16
		By sending an SMS	A7, A13, A9, A14, A15, A16
6		By personal knowledge	A16
7	By internet	By Google search	A1, A2, A3, A17
		By website search	A1, A5, A3, A6*, A17
		By visiting a web forum	A1, A5, A6*, A17
8	By personal contact	By calling a friend	A7, A8, A9, A10, A11, A12, A17
		By sending an SMS	A7, A13, A9, A14, A15, A17
9		By personal knowledge	A17
10	By internet	By Google search	A1, A2, A3, A18
		By website search	A1, A5, A3, A6*, A18
		By visiting a web forum	A1, A5, A6*, A18
11	By personal contact	By calling a friend	A7, A8, A9, A10, A11, A12, A18
		By sending an SMS	A7, A13, A9, A14, A15, A18
12		By personal knowledge	A18
13		By internet shopping	A19, A20
14		By physical shopping	A21
15		By failure	A22

Strategies discovery

We recorded 90 traces of activities produced by 90 students for which we know the sequence of selected strategies. As mentioned earlier, the knowledge of the strategies allows us working with the framework of Supervised Map Miner Method to estimate the parameters of the HMM, i.e. the coefficients of the matrices S and A . Since the intentional model comprises 15 strategies and 22 activities, the size of S is 22×22 and the size of A is 15×22 .

The emission matrix \hat{E} , shown in Table 7, provides some information on the students' behavior analysis when buying a present.

- One ascertainment is that A_1 : *open a browser*, A_2 : *type a query* and A_6 : *navigate on the website* activities are performed for the strategies S_1 , S_4 , S_7 and S_{10} equally with probabilities of 23%, 9 % and 15%, respectively (second, third and seventh row in table 7). This is not a surprising behavior since the chosen strategies are: by internet, consequently, the students first opened the browser, typed the query and then navigated on the website.

- The A_4 activity: *create idea concept* appears with a high probability of 43% (fifth row, third column). This observation is interesting because it shows the students tend to ask directly the friend for whom they would like buy the present.

Table 7. The emission matrix of obtained intentional model.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15
A1	0.23	0	0	0.23	0	0	0.23	0	0	0.23	0	0	0	0	0
A2	0.09	0	0	0.09	0	0	0.09	0	0	0.09	0	0	0	0	0
A3	0	0	0	0.15	0	0	0.15	0	0	0.15	0	0	0	0	0
A4	0.15	0.43	1	0	0	0	0	0	0	0	0	0	0	0	0
A5	0.23	0	0	0.15	0	0	0.15	0	0	0.15	0	0	0	0	0
A6	0.15	0	0	0.15	0	0	0.15	0	0	0.15	0	0	0	0	0
A7	0.15	0.11	0	0	0.10	0	0	0.16	0	0	0.16	0	0	0	0
A8	0	0.06	0	0	0.05	0	0	0.07	0	0	0.07	0	0	0	0
A9	0	0.11	0	0	0.10	0	0	0.16	0	0	0.16	0	0	0	0
A10	0	0.06	0	0	0.05	0	0	0.07	0	0	0.07	0	0	0	0
A11	0	0.05	0	0	0.05	0	0	0.07	0	0	0.07	0	0	0	0
A12	0	0.06	0	0	0.05	0	0	0.07	0	0	0.07	0	0	0	0
A13	0	0.06	0	0	0.05	0	0	0.08	0	0	0.08	0	0	0	0
A14	0	0.06	0	0	0.06	0	0	0.08	0	0	0.08	0	0	0	0
A15	0	0	0	0	0.06	0	0	0.08	0	0	0.08	0	0	0	0
A16	0	0	0	0.23	0.43	1	0	0	0	0	0	0	0	0	0
A17	0	0	0	0	0	0	0.23	0.16	1	0	0	0	0	0	0
A18	0	0	0	0	0	0	0	0	0.23	0.16	1	0	0	0	0
A19	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0
A20	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0
A21	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
A22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- The apparition of A_{16} : *modify an idea concept* with a high probability of 23%, 43% and 100% for strategies S_4 , S_5 and S_6 , respectively, indicates the percentages of the students who asked directly their friends and then changed their present idea. In the same way, the students who have asked directly their friend modified the place to buy a present.

- The activity A_{22} : *give up* has never been performed and consequently strategy S_{15} has never been chosen by the students. This observation means all students have successfully finished their shopping using prescribed Map guidance.

Intentional model discovery

Figure 9 illustrates the obtained intentional model using the best estimated parameter of HMM, i.e. the transition matrix, obtained in the previous step. This intentional model shows all the transitions between the strategies and the intentions. This map is obtained with a threshold adjusted to $\epsilon=0.1$. The three intentions *Find an idea to buy a present*, *Find a place to buy the present* are identified by the potatoes and *Stop* by the oval.

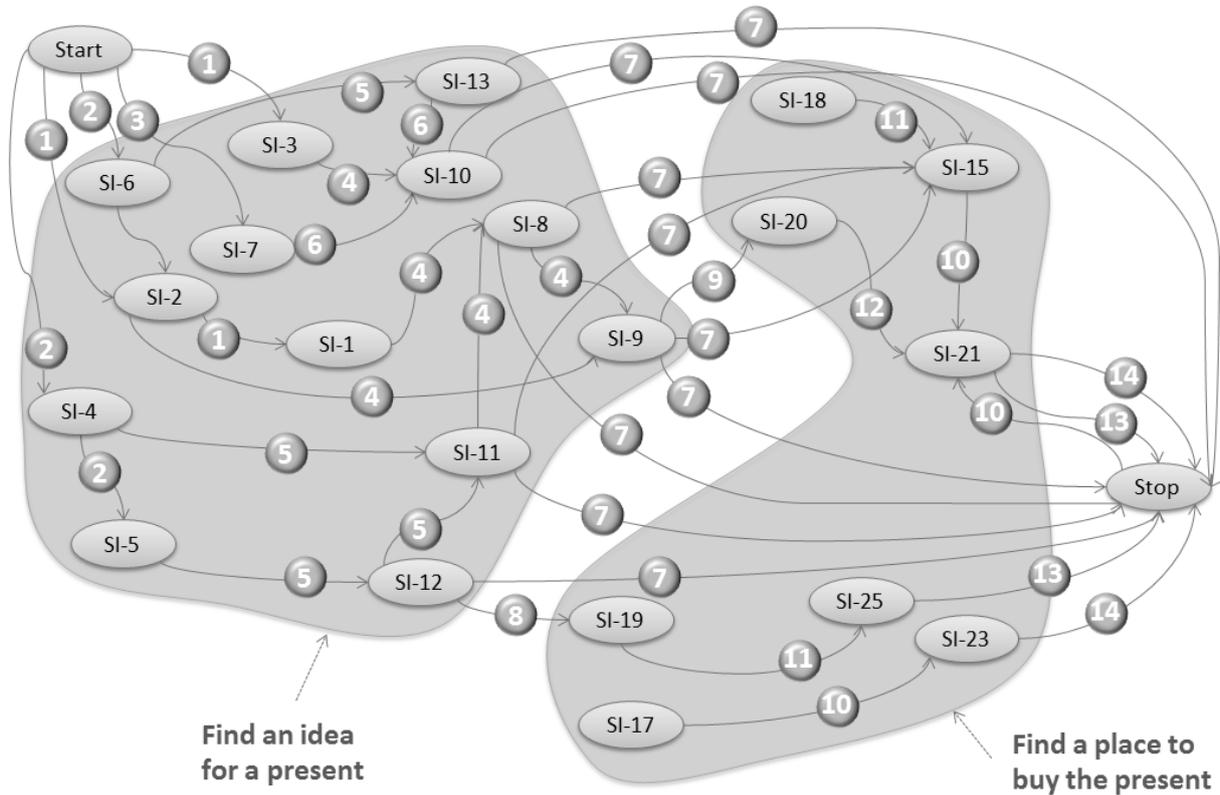


Figure 9. The obtained intentional model using the Supervised Map Miner Method with a threshold adjusted to $\varepsilon = 0.1$.

Figure 10 shows the high-level of abstraction of the obtained intentional model. This map provides some information about the students' behaviors:

- We observe that the prescribed map has been accurately followed by the students most of the time. Indeed, the transitions between the strategies we obtain with the Supervised Map Miner Method are mostly consistent with the prescribed map.
- However, we observe some deviations from the prescribed map. For instance, all the sections comprising strategies S_7 and S_8 were not chosen by students, i.e. *by visiting a web forum* and *by calling a friend*, respectively. We then only let S_{7a} as *by Google search* and S_{7b} as *by website search*, and S_8 as *by sending a SMS* in Figure 10.
- The students never selected the strategy 'by failure' to quit (strategy S_{15}) between the intentions *Find a place to buy the present* and *Stop*. This behavior means all students who have succeeded to find a place to buy the present, bought it either by internet shopping or by physical shopping.
- The students chose strategy S_7 *by Google search* and *by website search* more than strategies S_8 and S_9 to reach the intention *Find a place to buy the present*. This observation means they prefer to search for a place to buy a present on the internet rather than using personal contact or personal knowledge.

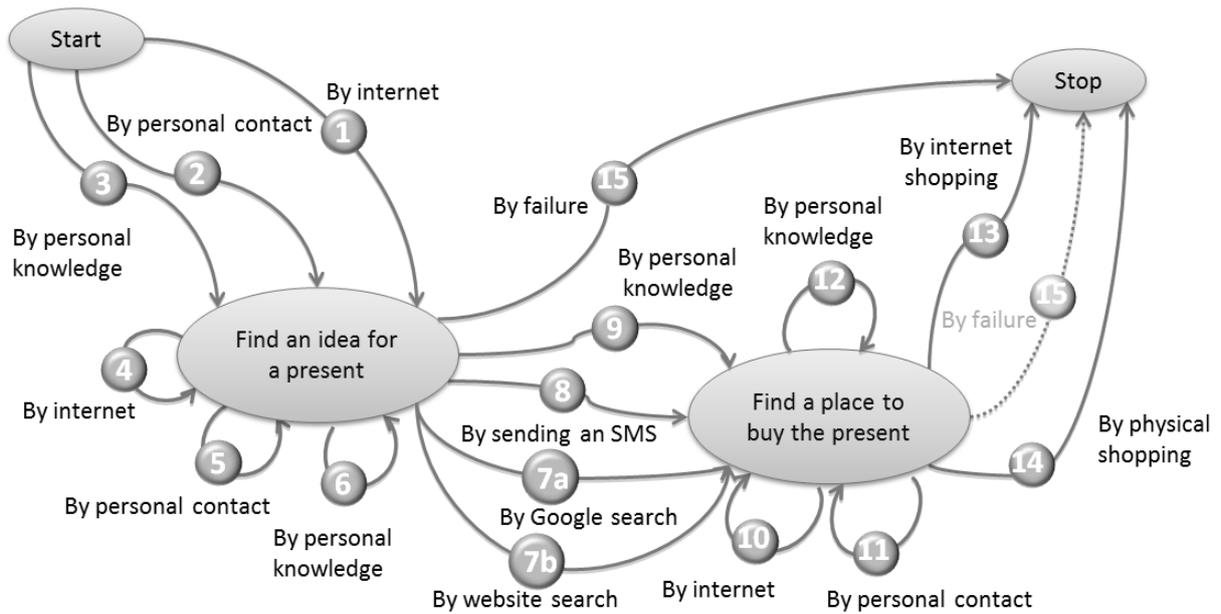


Figure 10. The high-level of abstraction for obtained map.

In this second controlled experiment, we verified our hypothesis, i.e. the impact of having more strategies in the prescribed map avoid deviations, and the deviations are less frequent than in the first experiment.

Results analysis

We calculate the recall, precision and F-score for the fifteen strategies. We only show the results of three strategies S_1 , S_8 and S_{12} for which these measures are not near to 100%. Figure 11 depicts these measures averaged over 90 test sequences for the strategies S_1 , S_8 and S_{12} .

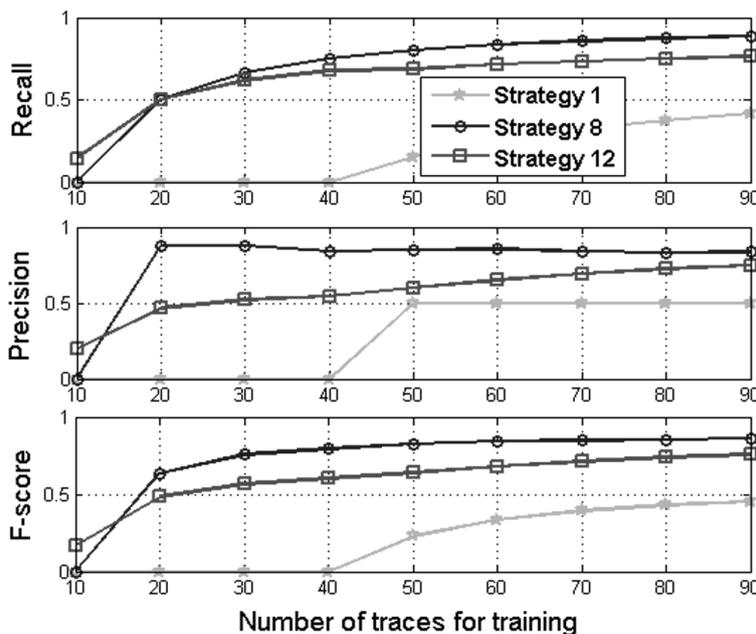


Figure 11. The results of Recall, Precision and F-score for strategies 1, 8 and 12.

The recall curve for strategy S_8 is stabilized around 89% which means the VA finds 89% of activities related to the strategy S_8 . In other words, a huge part of activities associated to strategy S_8 are identified with the VA. It means that for a length of 66, the VA provides stable results. The results of recall are presented for the three selected strategies in Table 8. It indicates the recall measure for strategy S_1 is of 41% which is not satisfying. The weak result of recall can be due to the fact that the estimation sequence is not long enough to detect all the behaviors and the VA needs more training samples to learn.

Table 8. Recall, precision and F-score for 3 strategies.

	Recall	Precision	F-SCORE
Strategy 1	41%	50%	0.45
Strategy 8	89%	83%	0.87
Strategy 12	77%	75%	0.75

The precision result stabilizes at 50% for strategy 1 which means 50% of the activities are associated to strategy S_1 while they belong to other strategies. The precision give better result for strategy 8 and 12.

Finally, the results of F-score express the reliability of the VA to detect the related strategies sequence to any activities sequence. For example, if the strategy S_{12} is chosen by a given student, the VA is able to find this strategy with an accuracy of 0.75%.

The mean values of recall, precision and F-score over all strategies which express the global performance of traces estimation are illustrated in Figure 12. The recall curve starts from 0.42 and its value increases until reaching the maximum value at 0.60. The curve of precision starts from 0.40 and stops at 0.71. This means the accuracy of retrieval for fifteen strategies is 0.6 which is more than satisfying.

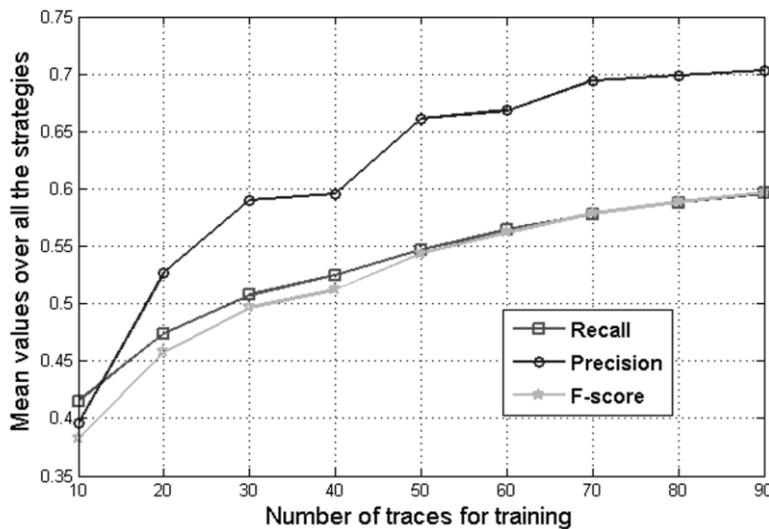


Figure 12. The results of mean values for recall, precision and F-score over all the strategies.

Threats to Validity

There are several threats to validity that may have impacted our work. One threat is related to the supervised learning assumption, i.e. the prescribed map is actually followed by students. This assumption is necessary to proceed with the discovery phase to allow VA assigning related strategies to the new activities sequence. However, during the enactment of the

process, students might have deliberately or by accident not followed the intentional model; they could also have poorly recorded their traces as shown in the sections 4.1.c and 4.2.c. Consequently, assuming that the prescribed model is followed by the students, this creates a bias in the definition of strategies and intentions.

Another threat is the absence of ground truth for labeling the activities sequences. Consequently, the labeling could be affected as it is a subjective process. Further, assigning the labels to the strategies and intentions constrains the discovered map to a limited space which leads to poor performance of supervised learning. In addition, supervised learning is a time-consuming and intensive labor due to human's effort required to label the data. The cost of labeling the data for supervised learning is quite high as it involves the users' commitment to label and comment their activities at each step of the process enactment.

CONCLUSION

Intention mining is a promising field of research with multiple applications both inside and outside the context of IS Engineering.

This paper makes two contributions: (a) an original intention mining method, called Supervised Map Miner Method, that generates intentional models formalized with the Map metamodel from traces of activities, and (b) a validation performed based on two controlled experiments; the validation evaluates the precision, recall and F-score of the results obtained with the Supervised Map Miner Method.

The Supervised Map Miner Method can be distinguished from the other approaches of intention mining field by several novelties: (a) the inputs in our approach are set of traces of activities related to users and no single entity as queries, logs, etc. (b) contrary to the other approaches, our method analyzes processes enacted by users to obtain a common goal, (c) the multi-level topology of activities, strategies and intentions requires a mathematical model, whereas in the other approaches, using classification techniques is enough to classify a single input into a class of intention, (d) our method offers the high-level intentions as the outputs of process model, contrary to the other approaches that discover only the low-level intentions. This difference is fed by the fact that we assume there are strong correlation and dependency between users' activities and they are not a single, independent and uncorrelated entity.

In practice, the Supervised Map Miner Method could be used whenever intentions are known and can be modeled in advance. This is for instance the case in the Enterprise Architecture context where not only business process are supposed to be implemented by their organizations, but also strategic objectives, missions, business goals are well-known and can be modeled (Thévenet & Salinesi, 2007). In this situation, the Supervised Map Miner Method could be used to monitor alignment with best practice strategies (for audit), target strategies (to monitor change), or to facilitate strategic emergence. In the method engineering context, the Supervised Map Miner Method can be used to understand which methods are actually used and how (Janković et al., 2013), and to inform recommendation techniques for people who need guidance (Epure, 2013) (Epure et al., 2014). There is little doubt many other applications will be found in the future.

Three key questions shape our future works. The first two ones are, what if intentions are not known and modeled in advance, and, how to mine to other intentional formalisms than Map? Both issues are hard: unsupervised learning is an obvious answer to the first one. This approach is promising as it generally shows better results than supervised learning. However, the problem is that we would like to keep making mined intentions explicit. This implies that a new element, such as ontology, should be introduced in the method to generate the names of the intentions. Our approach should support different formalisms to provide models that really match the users knowledge and organizations and projects culture, as I* and KAOS that are

widely used by the communities for goal modelling. We could try to abstract the intentions and strategies from the specificities of map models and generalize them to become useful for any kind of goal modeling approach. A complex framework like i* will be difficult to handle unless a solution is found to match all kinds of dependencies with observations. Our last issue is more difficult to handle: would it be possible to differentiate between intentional and non-intentional activities? A first step would be to undertake experiments where traces include activities that are not all instance of processes, but also random, non-rationale, or mechanistic.

Acknowledgment

We would like to thank Sébastien Sion for the development of the first version of the web-based tool, and the Master and Bachelor Miage students for participating in the experiments.

References

- Ajzen, I., & Fishbein, M. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*, Addison-Wesley, Reading, MA.
- Amyot, D., Horkoff, J., Gross, D., & Mussbacher, G. (2009). A lightweight GRL profile for i* modeling. In C. A. Heuser, G. Pernul (Eds.), *Proceedings of the ER 2009 Workshops, Advances in Conceptual Modeling - Challenging Perspectives* (pp. 254-264). Berlin Heidelberg: Springer.
- Assar, S., Ben Achour, C. & Si-Said, S. (2000). *Un Modèle pour la spécification des processus d'analyse des Systèmes d'Information*. In Proceedings of 18ème Congrès INFORSID (pp. 287-301).
- Baeza-Yates, R., Calderón-Benavides, R. & González-Caro, C. (2006). The intention behind web queries. In F. Crestani, P. Ferragina & M. Sanderson (Eds.), *Proceedings of the 13th International Conference on String Processing and Information Retrieval* (pp. 98-109). Berlin, Heidelberg: Springer.
- Bonito da Silva Santos, L.O., Guizzardi, G., Pires, L.F. & Van Sinderen M., (2009). From User Goals to Service Discovery and Composition. In C.A. Heuser & G. Pernul (Eds), *Proceedings of ER Workshops on Advances in Conceptual Modeling - Challenging Perspectives* (pp 265-274). Berlin, Heidelberg: Springer-Verlag.
- Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press.
- Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., & Mylopoulos, J. (2004). Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8 (3), 203-236.
- Brinkkemper, S. (1996). Method Engineering: Engineering of Information Systems Development Methods and Tools. *Information and Software Technology*, 38 (4), 275-280.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum* 36 (2), 3-10.
- Cambridge Advanced Learner's Dictionary & Thesaurus (2011)
<http://dictionary.cambridge.org/dictionary/british/>
- Christie, B. (1981). *Face to File Communication: A Psychological Approach to Information Systems*. New-York, USA: Wiley.
- Chulef, A.S., Read, S.J. & Walsh, D.A. (2001). A Hierarchical Taxonomy of Human Goals. *Motivation and Emotion Journal*, 25 (3), 191-232.
- Chung, C.C. & Lin, C.-J. (2001). *LIBSVM: A Library for Support Vector Machines* (Technical Report). Department of Computer Science and Information Engineering, National Taiwan University, Taiwan.
- Clauzel, D., Sehaba, K., & Prié, Y. (2009). Modelling and Visualising Traces for Reflexivity in Synchronous Collaborative Systems. In Y. Badr, Y., S. Caballé, S., F. Xhafa, F., A.

- Abraham, A., & B. Gros (Eds.), *Proceedings of the 1st International Conference on Intelligent Networking and Collaborative Systems* (pp. 16-23). IEEE.
- Dardenne A., van Lamsweerde A., Fickasu S. (1993). Goal-directed requirements acquisition. *Science of Computer Programming*, 20 (1-2), 3-50.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35 (8), 982-1003.
- Epure, E.V. (2013). *Intention-mining: A solution to process participant support in process aware IS* (Master thesis). Utrecht University, The Netherlands.
- Epure, E.V., Hug, C., Deneckère, R. & Brinkkemper, S. (2014). What Shall I Do Next? Intention Mining for Flexible Process Enactment. *26th International Conference on Advanced Information Systems Engineering*, Thessaloniki, Greece.
- Forney, G. D., Jr (1973). *The viterbi algorithm*. *Proceedings of the IEEE*, 61 (3), 268-278.
- Gales, M. J. (1998). Maximum likelihood linear transformations for HMM-based speech recognition. *Computer speech & language*, 12 (2), 75-98.
- González-Caro, C. & Baeza-Yates, R. (2011). A multi-faceted approach to query intent classification. In R. Grossi, F. Sebastiani & F. Silvestri (Eds.), *Proceedings of the 18th International Conference on String Processing and Information Retrieval* (pp. 368-379). Berlin, Heidelberg: Springer.
- Hashemi, R.R., Bahrami, A., LaPlant, J. & Thurber, K. (2008). Discovery of Intent through the Analysis of Visited Sites. In H.A Arabnia & R.R. Hashemi (Eds.), *Proceedings of the 2008 International Conference on Information & Knowledge Engineering* (pp. 417-422). CSREA Press.
- Jackson, M. (1995). *Software Requirements & Specifications, a Lexicon of Practice, Principles and Prejudices*. ACM Press, Addison-Wesley.
- Janković, M., Bajec, M., Khodabandelou, G., Deneckère, R., Hug, C. & Salinesi C. (2013). Intelligent Agile Method Framework. In L. Maciaszek & J. Filipe (Eds), *Proceedings of 8th International Conference on Evaluation of Novel Approaches to Software Engineering* (pp. 187-192). SCITEPRESS.
- Jarke, M., Pohl, K., Rolland, C. & Schmitt, J.R. (1994). Experienced-Based Method Evaluation and Improvement: A Process Modeling approach. In A.A. Verrijn-Stuart & T.W. Olle (Eds.), *Methods and Associated Tools for the Information Systems Life Cycle, Proceedings of the IFIP WG8.1 Working Conference* (pp. 1-27). The Netherlands: Elsevier, IFIP Transactions.
- Jaufman, O., Dold, A., Haeberlein, T., Schlumpberger, C. & Stupperich, M. (2004). Requirements for flexible software development processes within large and long taking projects. In *Proceedings of the 5th Meeting for Quality in Information and Communications Technology*, Portugal.
- Jethava, V., Calderón-Benavides, L., Baeza-Yates, R., Bhattacharyya, C. & Dubhashi, D. (2011). Scalable Multi-Dimensional User Intent Identification using Tree Structured Distributions. In W.-Y. Ma, J.-Y. Nie, R.A. Baeza-Yates, T.-S. Chua & W.B. Croft (Eds.), *Proceedings of the 34th International ACM Conference on Research and development in Information Retrieval* (pp. 395-404). New York, NY, USA: ACM.
- Kaabi R.S. & Souveyet C. (2007). Capturing Intentional Services with Business Process Maps. In C. Rolland, O. Pastor & J.-L. Cavarero (Eds.), *Proceedings of the 1st International conference in Research Challenges in Information Science* (pp 309-318). IEEE.
- Khodabandelou, G., Hug, C., Deneckère, R. & Salinesi, C. (2013). Process Mining versus. Intention Mining. In S. Nurcan, H.A. Proper, P. Soffer, J. Krogstie, R. Schmidt, T. Halpin & I. Bider (Eds), *Enterprise, Business-Process and Information Systems Modeling*, 147, Lecture Notes in Business Information Processing, Berlin Heidelberg: Springer, 466-480.

- Khodabandelou, G., Hug, C., Deneckère, R. & Salinesi, C. (2013). Supervised Intentional Models Discovery using Hidden Markov Models. In R. Wieringa, S. Nurcan, C. Rolland & J.-L. Cavarero (Eds.), *Proceedings of Seventh International Conference on Research Challenges in Information Science* (pp.1-11). IEEE.
- Kröll, M. & Strohmaier, M. (2009). Analyzing Human Intentions in Natural Language Text. In Y. Gil & N. Fridman Noy (Eds.), *Proceedings of the 5th International Conference on Knowledge Capture* (pp. 197-198). New York, NY, USA: ACM.
- Laflaquière, J., Lotfi, Settouti, S., Prié, Y. & Mille, A. (2006). Trace-Based framework for experience management and engineering. In B. Gabrys, R.J. Howlett & L.C. Jain (Eds.), *Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, 1* (1), Berlin, Heidelberg: Springer, 1171-1178.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.
- Liu, H. & Singh, P. (2004). ConceptNet — a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22 (4) 211-226.
- Martin, J. J. (1967). *Bayesian decision problems and Markov chains*. Wiley.
- Mei, T., Hua, X.-S. & Zhou, H.-Q. (2005). Tracking users' capture intention: a novel complementary view for home video content analysis. In *Proceedings of the 13th annual ACM International Conference on Multimedia* (pp. 531-534). New York, NY, USA: ACM.
- Merriam-Webster Dictionary. Merriam-Webster (2002)
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1990). WordNet: An online lexical database. *International Journal of Lexicography*, 3 (4) 235-244.
- Mirbel, I. & de Rivieres, V. (2002). Adapting Analysis and Design to Software Context: The Jecko Approach. In Z. Bellahsene, D. Patel & R. Rolland (Eds.), *Proceedings of the 8th International Conference on Object-Oriented. Information Systems* (pp. 223-228). Lecture Notes in Computer Science, Springer.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A. (2012). *Foundations of Machine Learning*. The MIT Press.
- Nurcan, S. (2008, January). A survey on the flexibility requirements related to business processes and modeling artifacts. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (pp. 378-378), IEEE.
- Nurcan, S., Etien, A., Kaabi, R., Zoukar, I., & Rolland, C. (2005). A strategy driven business process modelling approach. *Business Process Management Journal*, 11 (6), 628-649.
- Murphy, K. P. (2002). *Dynamic bayesian networks: representation, inference and learn* (PhD thesis). University of California.
- Object Management Group (2011). Business Process Model and Notation. <http://www.omg.org/spec/BPMN/2.0/>
- Outmazgin, N. & Soffer, P. (2013). Business Process Workarounds: What Can and Cannot Be Detected by Process Mining. *Lecture Notes in Business Information Processing*, 147, 48-62.
- Park, K., Lee, T., Jung, S., Lim, H. & Nam, S. (2010). Extracting Search Intentions from Web Search Logs. In *2nd International Conference on Information Technology Convergence and Services* (pp. 1-6). NY, USA: IEEE.
- Pérez-Castillo, R., Weber, B., García Rodríguez de Guzmán, I. & Piattini, M. (2011). Process mining through dynamic analysis for modernising legacy systems. *IET Software*, 5 (3) 304-319.
- Prat, N. (1997, June). *Goal formalisation and classification for requirements engineering*. Paper presented at International Workshop on Requirements Engineering Foundations for Software Quality (REFSQ), Barcelona, Spain.

- Rabiner, L.R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In A. Waibel & K.F. Lee (Eds.). *Readings in speech recognition*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 267-296.
- Ralyté, J. & Rolland, C. (2001). An Assembly Process Model for Method Engineering. In K.R. Dittrich, A. Geppert & M.C. Norrie (Eds.) *Proceedings of the 13th International Conference on Advanced Information Systems Engineering*. London, UK: Springer-Verlag, 267-283.
- Rolland, C., & Salinesi, C. (2005). Modeling goals and reasoning with them. In C. Wohlin, A. Aurum (Eds), *Engineering and Managing Software Requirements* (pp. 189-217) Berlin Heidelberg: Springer.
- Rolland, C., (1993), Modeling the requirements engineering process. In *Information Modelling and Knowledge Bases V: Principles and Formal Techniques: Results of the 3rd European-Japanese Seminar*, (pp.85-96), Budapest, Hungary.
- Rolland, C., Ben Achour, C., Cauvet, C., Ralyté, J., Sutcliffe, A., Maiden, N., Jarke, M., Haumer, P., Pohl, K., Dubois, E. & Heymans P. (1998). A proposal for a scenario classification framework. *Requirements Engineering*, 3 (1) 23-47.
- Rolland, C., Prakash, N. & Benjamin, A. (1999). A Multi-Model View of Process Modeling. *Requirements Engineering*, 4 (4) 169-187.
- Rozinat, A. (2010). *Process mining: conformance and extension* (PhD thesis). Eindhoven Technische Universiteit.
- Rumelhart, K. E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro (Ed), *Theoretical Issues in Reading Comprehension*. Hillsdale: Lawrence Erlbaum.
- Sadikov, E., Madhavan, J., Wang, L. & Halevy, A. (2010). Clustering Query Refinements by User Intent. In M. Rappa, P. Jones, J. Freire & S. Chakrabarti (Eds.), *Proceedings of the 19th International Conference on World Wide Web* (pp. 841-850). New York, NY, USA: ACM.
- Schein, A., Popescul, A. & Ungar, L. (2001). *Pennaspect: A two-way aspect model implementation* (Technical report). The University of Pennsylvania, Department of Computer and Information Science.
- Standish group, CHAOS Report, 2011.
- Strohmaier, M. & Kröll, M. (2012). Acquiring knowledge about human goals from Search Query Logs. *Information Processing & Management*, 48 (1) 63-82.
- Swanson, E. B. (1974). Management information systems: appreciation and involvement. *Management Science*, 21 (2), 178-188.
- Thevenet, L.H, & Salinesi, C. (2007). Aligning IS to Organisations' Strategy: the INSTAL Method. In J. Krogstie, G. Sindre & A. Opdahl (Eds.), *Proceedings of the 19th International Conference on Advanced Information Systems Engineering*, Berlin, Heidelberg: Springer-Verlag, 203-217.
- Tiwari, A., Turner, C. J., & Majeed, B. (2008). A review of business process mining: state-of-the-art and future trends. *Business Process Management Journal*, 14 (1), 5-22.
- van der Aalst, W.M.P. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin, Heidelberg: Springer.
- Van der Aalst, W. M., & Weijters, A. J. M. M. (2004). Process mining: a research agenda. *Computers in Industry*, 53 (3), 231-244.
- Witten, I. H. & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, 2nd edition.
- Yu, E. S. K. (1995). *Modelling Strategic Relationships for Process Reengineering* (PhD thesis). University of Toronto, Department of Computer Science.

Yu, E. S. K., & Mylopoulos, J. (1994, May). Understanding “why” in software process modelling, analysis, and design. *Proceedings of the 16th International Conference on Software Engineering* (pp. 159-168). IEEE Computer Society Press.

Zheng, C., Fan, L., Huan, L., Yin, L., Wei-Ying, M. & Liu, W. (2002, November). User Intentional modeling in Web Applications Using Data Mining. *World Wide Web*, 5 (3) 181-191.