Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 37 (2014) 203 – 210

Procedia Computer Science

The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)

A Semantic Rule-Based Approach Supported by Process Mining for Personalised Adaptive Learning

Kingsley Okoye[a], Abdel-Rahman H. Tawil[a], Usman Naeem[a], Rabih Bashroush[a], Elyes Lamine[b]

[a]*School of Architecture Computing and Engineering, University of East London, London, UK*

[b]*Université de Toulouse, Mines-Albi, CGI, Campus Jarlard, Albi Cedex 09, France*

Abstract

Currently, automated learning systems are widely used for educational and training purposes within various organisations including, schools, universities and further education centres. There has been a big gap between the extraction of useful patterns from data sources to knowledge, as it is crucial that data is made valid, novel, potentially useful and understandable. To meet the needs of intended users, there is requirement for learning systems to embody technologies that support learners in achieving their learning goals and this process don't happen automatically. This paper propose a novel approach for automated learning that is capable of detecting changing trends in learning behaviours and abilities through the use of process mining techniques. The goal is to discover user interaction patterns within learning processes, and respond by making decisions based on adaptive rules centred on captured user profiles. The approach applies semantic annotation of activity logs within the learning process in order to discover patterns automatically by means of semantic reasoning. Therefore, our proposed approach is grounded on Semantic Modelling and Process Mining techniques. To this end, it is possible to apply effective reasoning methods to make inferences over a Learning Process Knowledge-Base that leads to automated discovery of learning patterns or behaviour.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.

***Keywords**:* process model, semantic rules, process mining, user profile, learning behaviour, event logs

* Corresponding author. Tel.: +44-78-6194-2421; fax: +44-20-8223-2963.
E-mail address: u0926644@uel.ac.uk

1877-0509 © 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Peer-review under responsibility of the Program Chairs of EUSPN-2014 and ICTH 2014.
doi:10.1016/j.procs.2014.08.031

1. Introduction

Privation of conformance and suitability of automated learning contents is having an increasingly debilitating impact on learning, which has a strong influence on expected learning outcomes [1]. Studies have shown that challenges in current information-rich world is not only to make information available for learners at any time or in any form, but should essentially offer the right content to the right user and in the right format [2, 3]. This is necessary to provide continuous intelligent recommendation of leaning patterns and feedback on learner's performance. The conceptual knowledge of monitored learning behaviour can be semantically annotated to adaptively suggest appropriate future learning paths, using user's profile information to improve the learning progression [4]. Events within a process can be related to exactly one case and assigned a case identifier [5] which results in automatic creation of workflow processes achieved by using a semantic annotation scheme to represent the event logs about the user.

In this paper, process mining is used to discover sets of recurrent behaviours that can be found within a learning process. As a result, suitable learning patterns are determined by means of semantic reasoning which can then be used to address the problem of adapting learning to the captured user profiles.

The rest of the paper is structured as follows; in Section 2, appropriate related work is analysed and discussed. Section 3, presents our proposed learning model to express user profiles as well as the representation of learning components. In addition, we discuss the generation of the learning process model and how we semantically annotate the process model describing in detail its ontological representation and reasoning using Ontology Web languages. Section 4 shows our approach to automated discovery of learning behaviours from the analysis of event logs and Actions. The prototype implementation and preliminary outcomes are discussed in Section 5. Finally, Section 6 concludes the paper and points out directions for future research.

2. Related Work

In recent years, several researches have been focused on applying process mining technologies to different aspects of business processes. Most of the developed systems use process mining techniques only for representation of business concepts, knowledge or data [5]. In our approach, we utilize process mining techniques to represent learning processes. Our focus is to further enhance this area of research by not only adapting the process mining tools but also present a way to relate semantic-based reasoning for adaptation within the learning process. Given a set of actions from event log of a learning process, our proposed approach automatically constructs process models capable of describing and enhancing observed behaviours.

Fahland and Van der Aalst [6] note that it is difficult to learn useful models from event logs following the characteristics of real-life events. However, process mining has been proved as one of the existing technologies that typically aim to extract non-trivial and useful information from event logs [3, 5]. Although, past study in these area argue that heuristic mining, genetic mining, and fuzzy mining [7] provide case-hardened process discovery techniques capable of constructing intuitive simple models to explain the most likely or common behaviours.

Process discovery, which lately has been seen as the most important and most visible intellectual challenge related to process mining aims to automatically construct a process model e.g., a Petri net or a BPMN model [5] and describes causal dependencies between activities [6]. In principle one could use process discovery to obtain a model that describes reality. The second type of process mining is conformance checking where, an existing process model is compared with an event log of the same process to check if in reality it conforms to the resulting model [8, 9]. Conformance check could imply that the model does not describe the executed process as observed in reality or is being executed in a different order. It could also mean that activities in the model are skipped in the log or that the log contains events not described by the model. Given this drawback, the last type of process mining; model enhancement comes into play. Van der Aalst et al [10] used the idea of an enhanced existing model to maintain compliance and to quantify deviations using information about the actual process recorded in some event logs from a business process.

Nooijen et al [11] recently introduced an automatic technique (arti-fact life cycle model) for discovering structured processes from a data source, by re-using a number of existing techniques to fill in crucial missing gaps in each notion of data object in Data-centric systems. In order to fill in these gaps, event types specification is utilized to construct database queries which extracts attributes from all event logs, groups them into cases, orders them by time stamps, and then writes the result into a classical logs in separate database columns. Many approaches have been tested to extract event logs from ERP (Enterprise Resource Planning) systems such as SAP [13] and PeopleSoft [14]. Consequently, as ERP systems in general provide multiple case identifiers, the majority of these approaches failed. The authors in [12] argue that success could only be reported when database tables are carefully selected by hand or a better view of data is semantically annotated.

Another research direction is focused on discovering common structures that can be found in a variety of processes describing the Workflow Activity Patterns (WAPS). Various definitions of workflow have been proposed in literature [15, 16]. Thom et al [16] describes WAPS as structures involving the interaction between the user and the control-flow constructs used to model the semantics of the activities being performed. Workflow systems assume that a process can be divided into small, unitary actions, called Activities [15]. To perform a given process, one must perform the set (or perhaps a subset) of the activities that comprise it. Hence, an Activity is an Action that is a semantic unit at some level, which can be thought of as a function that modifies the state of the process in terms of the semantics of the patterns and can be discovered automatically by means of semantic reasoning.

During the last decade workflow management concepts and technology have been applied in many enterprise information systems [17] such as Staffware, IBM, MQSeries, and COSA, which offer generic modelling and enactment capabilities for structured processes. Many other software systems have adopted workflow technology, for example ERP (Enterprise Resource Planning) systems such as SAP, PeopleSoft, Baan and Oracle, CRM (Customer Relationship Management) software. However, despite its advantages, many problems are still being encountered when applying workflow technology. One of the problems is that these systems require a workflow design [18]. This implies that a designer has to construct a detailed model accurately describing the routing of work, which most often requires deep knowledge of the workflow language and management involved.

Huang & Shiu [19] noted that searching for suitable learning paths and content for achieving a learning goal is time consuming and troublesome especially on dynamic learning platforms. To tackle these problems, the authors proposes a User-Centric Adaptive Learning System (UALS) that uses sequential pattern mining to construct adaptive learning paths based on users' collective intelligence and recorded events and then employs Item Response Theory (IRT) with collaborative voting approach to estimate learners' abilities for recommending adaptive materials.

Our research differs from these previous works in several aspects. First, we provide an automated learning process by means of semantic reasoning. We focus on personalising learning based on user's behaviour as opposed to most existing systems that provide guidance based on views of designers or experts. Second, we also support adaptive learning using process-mining techniques. Third, this work is not only intended to ensure learner's ability to learn or meet their learning needs but is expected to be useful in providing learning path and guidance based on individual differences. This is achieved by collecting user's initial capabilities and preferences on interaction using semantic rules and process mining technique to model and detect behavioural changes as well as determine which adaptations or further assistive measures are best suited or may be required through time.

3. Semantic Rule-Based Approach to Learning

Semantic Rule-based approach is expected to collect routines and monitor changes in user's behaviour during the learning process [19, 20], to determine which adaptations technique may be required progressively through time.

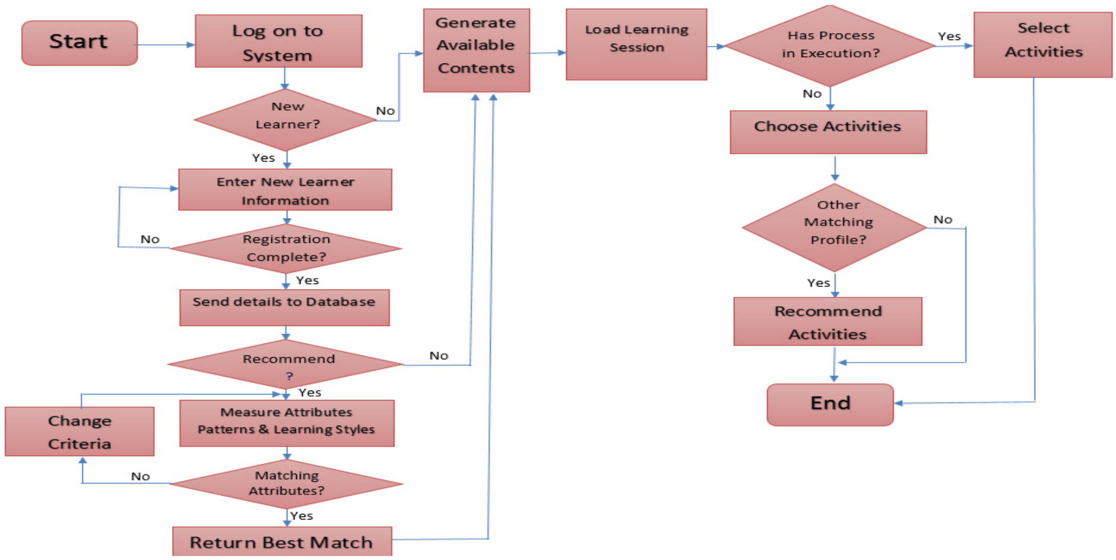

Fig. 1 Architectural Diagram of Process Workflow

In Fig 1. the semantic approach as illustrated, describes and takes into account users profile (prior knowledge of learners background), learning behaviour and actions when using the system.

3.1 Process Model

In this section, we utilise a rule mining approach to classify instances based on predictor variable to explain the dependent variables in terms of independent ones. Variants are a great way to compliment the way we look at processes. Captured data is simplified based on the variants by showing the processes in a more detailed way. Given the set of events within a learning process, Fig. 2 below represents a set of learning activities we processed using Disco Process Management System by Fluxicon [21].

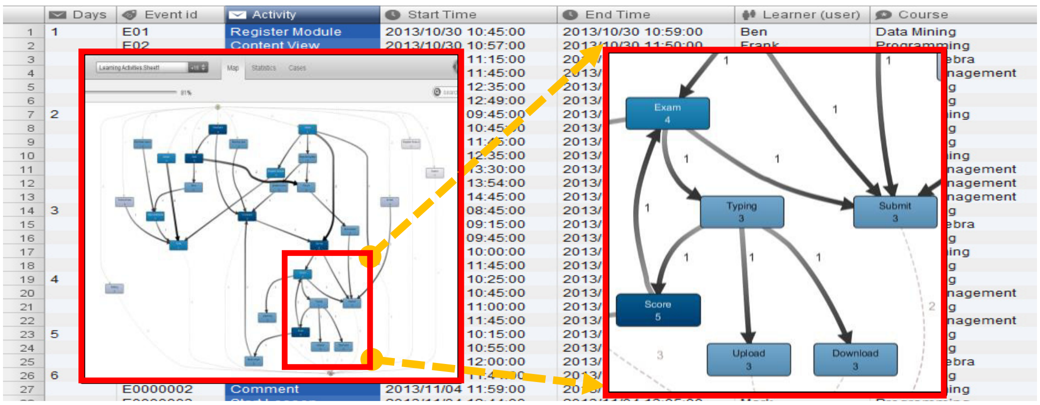

Fig. 2 Process in Execution: Learners log of activities processed using Disco

In Fig. 2, we imported the set of data into Disco to show in details how the processes have been performed. This approach describes our process mapping. The result is a mining approach that provides reliable and trustworthy results for data sets of arbitrary complexity and can be reasoned and understood efficiently by domain experts with no prior experience in process mining. The Disco miner is based on the proven framework of the Fuzzy Miner [7], but we developed in this paper a completely new set of process metrics and modelling strategies using a semantic rule-based approach. These additional metrics are very useful for process analysis because they hold relevant context information such as *Domain-specific characteristics.* We focus on the *Attribute filter* [21, 22] which describes as well as exclude certain activities, resources or process categories based on data attributes allowing us to quickly and interactively explore processes into multiple directions and to answer concrete questions about our process. In the next section, we describe and implement the concept of ontological reasoning of learning activities and actions capable of deducing inference to propose process model based on designed rule base that serves as a conceptual model for building our proposed semantic approach to automated learning.

3.2 Semantic Modelling and Reasoning

Ontology is the Science that can be used to model different kinds and structure of objects, events, processes and behaviours as they happen in reality, such as learning process as described in this paper. In this section, we describe our process model using the Web Ontology Language (OWL). Below is an OWL version 2 model for our learning process, which we implemented in Protégé 4.3 and reasoned upon using Pellet 2. Protégé OWL editor [23] supports Description Logic (DL) Queries and SWRL rules [26].

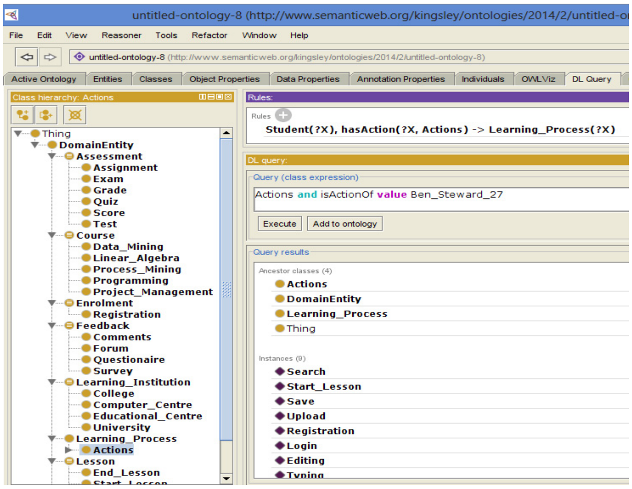

Fig. 3 Proposed learning Model in Protégé Editor

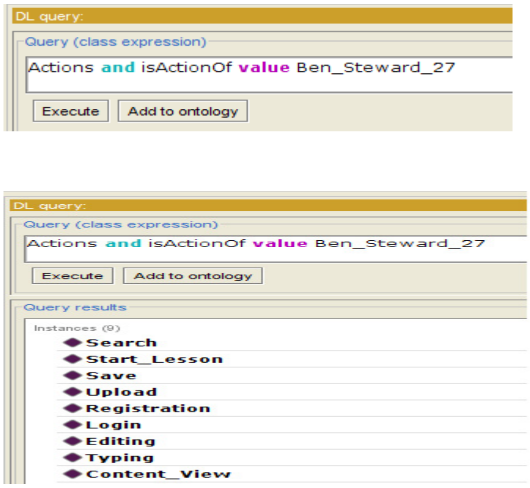

Fig 4. DL Query and the result of the query.

In Fig. 3, we used the protégé Editor to construct an ontology that expresses the functionality of our proposed model in terms of individual learning patterns. The *Cases* and *Actions* within the learning process were defined as sub-class of the main class *DomainEntity*. The *class expression* is based on the OWL syntax fundamentally focused on collecting all information about a particular class, property, or individual into a single construct, called a frame. The DL Query provides the platform for searching the classified ontology. In our model, DL Queries and SWRL rules were used to reason about OWL individuals, primarily in terms of OWL classes, Object Properties and Data Properties to infer the learning activities of any named individual. In Fig. 4, executing the query refer explicitly to OWL individual **Ben_Steward_27**. The result of the DL Query produces the instance value of Ben_Steward_27 Actions within the Learning Model. The result of the logic expression and reasoning of *Actions* for **Ben_Steward_27** is what we use to show our process model description and for automated discovery of learning patterns as implemented in subsequent sections.

4. Automated Discovery of Learning Patterns/Behaviour

Process Mining is one of the existing techniques that allows for traces not present in an existing process model to be discovered by using algorithms [5, 18] as well as rules to generalise and allow for behaviours unrelated to the ones in the log to be observed. The α-algorithm is one of the many algorithms used in process mining, aimed at reconstructing causality from sets of events sequence. It was first put forward by van der Aalst, Weijters and Măruşter [18]. Several extensions of it have since been presented, as we utilized below.

If the set of learning activities, as implemented in our OWL file (Fig. 3), are notated as follows; where **A = Enrolment, B = Lesson, C = Assessment and D = Feedback.** These Activities as they happen sequentially in reality take a workflow log $W \subseteq T^*$ as input and result in a workflow net being constructed. Workflow Logs $W \subseteq T^*$ is a definitive relationship management algorithm: where; W is a finite set of Events. T is a finite set of transitions such that W /T ¼; Hence, $W \subseteq T^*$ is a set of directed pattern, called the flow relation.

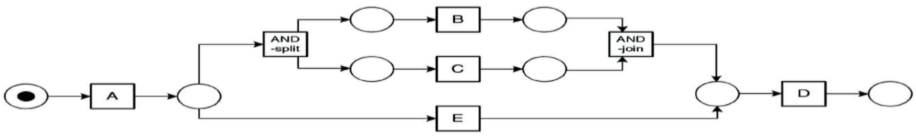

Fig 5. Control Flow of process model corresponding to the workflow log

In Fig. 5, it is easy to check that all the traces above are possible in the model. The initial marking (**A**) is enabled because of the token at the start of the learning activities. The control-flow for its execution (**X ⇒ Y**) results in the marking that learner (***X***) performs an activity that is casually followed by activity (***Y***) [46], hence, the stronger the relation between **X** & **Y**. Now this enables the execution of the remaining actions to the final event (**D**) in the path.

From our model in protégé, **user profile** (Ben_Steward_27) learning behaviour can be described by the sequence of his Actions in each case as follows;

Table 1. Sequence of Actions in a Learning Process

LearnerProfile: (Ben)		
Action(1)	=	***{Login}***
Action(2)	=	***{Registration}***
Action(3)	=	***{Search, Content_View}***
Action(4)	=	***{Typing, Editing}***
Action(5)	=	***{Save, Upload}***
Action(6)	=	***{Start_Lesson}***

Hence, the *α-algorithm* [18] for Ben learning behaviour discovered from the frequent sequence of the learning profile as described in Table 1. will be as follows:

(Action[1-6]) = {*Login, Registration, Search, Content_View, Typing, Editing, Save, Upload, Start_Lesson*}

These identified Actions can be modelled using the *Association Rule* in data mining [24, 25]. However, our approach aims at discovering similar rules but then without focusing on a particular variable to discover user interaction patterns and then respond by making decisions based on adaptive rules centred on the captured user profiles. The goal is to discover and create rules of the form;

X ⇒ Y (*IF* **X** *THEN* **Y)** where **X** = *Learning pattern* (Antecedent) and **Y**= *Learning pattern extension* (Consequent)

This rule is similar to the SWRL Syntax; *atom ^ atom .... → atom ^ atom* e.g. Learner(?X), hasAction(?X, ?Registration) -> Enrolment_Process(?X)

The approach makes it possible to efficiently generate learning patterns based on the sequence or control-flow of the learning behaviours. Driven by these variables from the Action logs, it is possible that the following learning path can be suggested to improve the semantics of learning patterns. *Rules* like "Learners that have similar instances as Student_Ben are most likely to come across Case(Ben)" can be derived.

Student(?X), isPartOf (?X, ?Course), hasActionSimilar (?Actions, "Ben") -> hasSimilarLearningPathTo(?X, "Ben")

Similarly, the *Association rule* states that when **X** occurs, then **Y** occurs with certain probability by using the frequent item-set (***I***) to generate rules [24]. This implies that for each process of non-empty Action (***A***) from our learning model in Fig. 3, $\mathbf{X} \Rightarrow \mathbf{Y}$ is an association rule of the form; $\mathbf{Y} = (\mathbf{X} - \boldsymbol{A})$ where, Confidence ($\mathbf{X} \Rightarrow \mathbf{Y}$) ≥ minimum Confidence

Support ($\mathbf{X} \Rightarrow \mathbf{Y}$) = support ($\mathbf{X} \cup \mathbf{Y}$) = support ($\boldsymbol{A}$). Hence, Confidence ($\mathbf{X} \Rightarrow \mathbf{Y}$) = support ($\mathbf{X} \cup \mathbf{Y}$) / support ($\boldsymbol{A}$)

Consequently, if **X** = **Case(Ben)** and **Y** = Ben**(Action[1-6])** and Suppose that **X** and **Y** are represented according to utilization factor of frequency ranging from 0.01 to 0.09, where **X** has a support of 0.04 and **Y** has a support of 0.03 from the represented variables; then the expression of certainty $\mathbf{X} \Rightarrow \mathbf{Y}$ is 0.03 / 0.04 = 0.75.

This means the prediction variable that 75% of Learners that has instances of **X** as in **Case(Ben)** may later come across **Y** as in **Ben (Action[1-6]).**

Hence, the Rule expression; Learner(?X), hasSimilarLearningPathTo(?X, "Ben") -> hasCase(?X, "Ben").

5. Discussion

The development of semantic mining tools entails three building blocks *Annotated Event Logs, Ontologies* and *Semantic Reasoning* that aim at discovering, conformance and extension of processes. Any pattern or learning behaviour can be discovered as a consequence or condition of a rule. Ontology can be layered on top of these existing information asset to provide more enhancements to real processes in the same manner as process mining. This characteristic is the ability to match same idea as well as use the coherence and structure itself to inform and answer questions of relationships. Hence, by specifying one concept (say *Learning_process*) one knows that we are also referring to another concept (*Learners or Students*) say *Learners* learn through engaging in a *Learning_process*.

The semantic approach described in this paper has been used to develop semantic process mining plug-ins. In this paper, we utilised an OWL version 2 model for our learning process, which we implemented using Protégé 4 and Pellet 2 [23] and Disco process management system by fluxicon [21] to process our data. In addition, the expressivity of our OWL was extended by adding SWRL rules to the implemented ontology based on the concrete syntax of the SWRL proposal [26]. This work has been able to utilise the frameworks as described in this paper to propose a semantic rule-based approach supported by process mining technique for personalised adaptive learning.

6. Conclusion and Future Work

In this paper, process mining is used to discover, monitor and improve the set of recurrent behaviours that can be found within learning processes. The technique is utilised in order to address the problem of determining the presence of different learning patterns in process models. A semantic process model such as the one described in this paper (User-Oriented Knowledge-Base System) will be of great impact and significance in this area to drive learning by using process mining techniques to discover rules through semantic reasoning and adopting web languages such as OWL and SWRL. The outcome is of great importance in bridging the gap between the levels of learning for different users by providing them with the same learning opportunity, through a system that adaptively support the personalisation of contents based on data regarding the users learning behaviour or Actions.

Future work will focus on first applying the approach described in this paper to a case study of users with a particular learning difficulty. Then, we intend to tap on the increase in recording of information related to study behaviour of individuals in many organisations that are involved in the educational sector. For this, we plan to use our technique to analyse study behaviour using a database containing detailed information about learning styles of students, as some of these educational processes are very unstructured The goal is to cover the whole spectrum of the approach presented in this paper to provide more authenticity and better support for automated learning systems.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments.

References

[1] Mancera L, Baldiris S, Fabregat R, Vinas F, Caparros B. Adapting Suitable Spaces in Learning Management Systems to Support Distance Learning in Adults with ADHD. *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies* (ICALT), 6-8th July 2011; pp. 105 – 109.
[2] Yu Z, Nakamura Y, Jang S, Kajita S, Mase K. Ontology-Based Semantic Recommendation for Context-Aware E-Learning. *Ubiquitous Intelligence and Computing,* 2011; vol. 4611, 2007, pp 898-907.
[3] IEEE Task Force on Process Mining. Process Mining Manifesto. *In: Business Process Management Workshops* LNBIP, Springer, 2012; vol. 99, pp. 169–194.
[4] Tawil ARH, Litnhouvongs A, Chevalier O, Taweel A. An Ontology-Based Approach to Adaptive Home Careflows. *In Proceedings of DEXA Workshops,* 2011; pp. 522-526.
[5] Van der Aalst WMP. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.
[6] Fahland D, Van der Aalst WMP. Repairing process models to reflect reality. *In Business Process Management 2012, vol. 7481 of Lecture Notes in Computer Science,* Springer, 2012; pp. 229–245.
[7] Gunther CW. Process Mining in Flexible Environments. *PhD Thesis, Eindhoven University of Technology.* September, 2009.
[8] Adriansyah A, Van Dongen B, Van der Aalst WMP. Conformance Checking using Cost-Based Fitness Analysis. *In proceedings of EDOC 2011*. IEEE Computer Society, 2011; pp. 55-64.
[9] Munoz-Gama J, Carmona J. Enhancing Precision in Process Conformance: Stability, Confidence and Severity. *In Proceedings of CIDM 2011*. IEEE, Paris, France, April, 2011.
[10] Van der Aalst WMP, Schonenberg MH, Song M. Time Prediction Based on Process Mining. *Journal of Information Systems*, 2011; Vol. 36(2), pp. 450–475.
[11] Nooijen EHJ, Van Dongen BF, Fahland D. Automatic Discovery of Data-Centric and Artifact-Centric Processes. *International Business Process Management (BPM) Workshops*. *Lecture Notes in Business Information Processing*, 2013; Vol. 132, pp. 316-327.
[12] Nganji JT, Brayshaw M, Tompsett B. Ontology-driven disability-aware e-learning personalisation with ONTODAPS. *Journal of Campus-Wide Information Systems*, 2013; Vol. 30(1), pp.17 – 34.
[13] Piessens D. Event Log Extraction from SAP. *In ECC 6*.0, 2011.
[14] Ramesh A. *Process mining*. In PeopleSoft. 2006.
[15] Ferreira DL, Thom LH. A Semantic Approach to the Discovery of Workflow Activity Patterns in Event Logs. *International Journal of Business Process Integration and Management*, Inderscience Publisher, 2012; Vol. 6(1).
[16] Thom LH, Reichert M, Iochpe C. Activity patterns in process-aware information systems: basic concepts and empirical evidence. *Journal of Business Process Integration and Management*, 2009; Vol. 4(2), pp. 93–110.
[17] Van der Aalst WMP, Van Hee KM. Workflow management: Models, Methods, and Systems. *MIT Press, Cambridge,* MA, 2004.
[18] Van der Aalst WMP, Weijters AJMM, Maruster L. Workflow Mining: Discovering Process Models from Event Logs. *International Journal of IEEE transactions on Knowledge and Data engineering,* 2004; Vol. 16(9). Pg. 1128-1142.
[19] Huang SL, Shiu JH. A User-Centric Adaptive Learning System for E-Learning 2.0. *International Journal of Educational Technology & Society*, 2012; Vol. 15 (3), pp. 214–225.
[20] Nganji JT, Brayshaw M, Tompsett B. Ontology-Based E-Learning Personalisation for Disabled Students in Higher Education. *Journal of ITALICS*, 2011; Vol. 10(1) pp. 1473-7507.
[21] Disco Software Download "http://fluxicon.com/disco/"
[22] Disco Users Guide. "https://fluxicon.com/disco/files/Disco-User-Guide.pdf"
[23] Protégé App. "http://protege.cim3.net"
[24] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques. [3rd Ed] *The Morgan Kaufmann Series in Data Management Systems,* Morgan Kaufmann Publishers, July 2011. ISBN 978-0123814791.
[25] Liu B. CS 583 Lecture *Data Mining and text mining* SEO 931 Department of Computer Science, University of Illinois at Chicago (UIC) Internet: http://www.cs.uic.edu/~liub/teach/cs583-fall-05/cs583.html, Aug. 20, 2005. Accessed [Feb. 18, 2014]
[26] SWRL: A Semantic Web Rule Language "http://www.w3.org/Submission/2004/SUBM-SWRL-20040521/"