Contributions of KDD to the Knowledge Management Process: a Case Study

Paulo de Tarso Costa de Sousa
Superior Electoral Court
SEPN 514 Ed Anexo II
55-61-340-3863
ptarso@tse.gov.br

Hércules Antonio do Prado
Brazilian Agricultural Research Corporation
Center for Agricultural Research on Savannah,
Caixa Postal 08.223 – CEP: 70770-000 - Planaltina, DF, Brasil
hercules@cpac.embrapa.br

Eduardo Amadeu Moresi
Catholic University of Brasília
Graduate Program in Management of Knowledge and Information Technology
CEP: 70.790-160 - Brasília, DF, Brasil
moresi@ucb.br

Marcelo Ladeira
University of Brasília
Campus Universitário Darcy Ribeiro - Asa Norte
CEP: 70910–900 – Brasília – DF - Brasil
mladeira@cic.unb.br

ABSTRACT

Knowledge Discovery in Databases (KDD), as any organizational process, is carried out beneath a Knowledge Management (KM) model adopted (even informally) by a corporation. KDD is grossly described in three steps: pre-processing, data mining, and post-processing. The latter is mainly related to the task of transforming in knowledge the patterns issued in the data mining step. On the other hand, KM comprises the following phases, in which knowledge is the subject of the actions: identification of abilities, acquisition, selection and validation, organization and storage, sharing, application, and creation. Although there are many overlaps between KDD and KM, one of them is broadly recognized: the point in which knowledge arises. This paper concerns a study aimed at clarifying relations between the overlapping areas of KDD and knowledge creation, in KM. The work is conducted by means of a case study using the data from the Electoral Court of the Federal District (ECFD), Brazil. The study was developed over a 1.717.000-citizens data set from which data mining models were built by applying algorithms from Weka. It was observed that, although the importance of Information Technology is well recognized in the KM realm, the techniques of KDD deserve a special place in the knowledge creation phase of KM. Moreover, beyond the overlap of post-processing and knowledge creation, other steps of KDD can contribute significantly to KM. An example is the fact that one important decision taken from the ECFD board was taken on the basis of a knowledge acquired from the pre-processing step of KDD.

Keywords

data mining, knowledge management, knowledge discovery in databases.

1. INTRODUCTION AND MOTIVATION

Practitioners on Knowledge Management (KM) usually adopt some model to guide the process and this model describes the activity of knowledge creation. On the other hand, the Knowledge Discovery in Databases (KDD) practice has been supported by methods like CRISP-DM [1], that include an interpretation activity. The clearest overlap of KDD and KM is the knowledge creation phase in KM and the interpretation task in KDD. In this paper we explore relations between KDD and KM phases by means of a case study in order to set a scene in which we show the specific importance of KDD to KM and try to make smoother the road of integration of both areas. It is shown how KDD can generate relevant input to KM not only from the data mining and post-processing tasks, but also from the pre-processing activities. Observations taken during tasks like problem definition, data acquisition and cleaning, and data and algorithm engineering can lead the analyst to be aware of some organizational failures. These failures, if not properly analyzed in a KM context, cannot receive the deserved importance, keeping the organization from benefiting from preventive or corrective actions. The discussion flows on the basis of an application on the Brazilian election domain, with focus on the organization data rather than in voting problems. CRISP-DM [1] method was applied to conduct the process while Weka [17] suite was adopted to build classification and clustering models. Specifically, a *k*-means algorithm was used for clustering and decision trees for classification. The reference model we used for KM is the generic one proposed by Stollenwerk [14] shown in Figure 1.

Figure 1 – Generic KM model of Stollenwerk

This model considers four aspects that compose the environment of KM: leadership, culture, technology, measures and compensations. Inside this environment, seven activities related to the generation of organizational knowledge are proposed. They

are: identification of abilities, knowledge capture, selection and validation, knowledge organization and storage, knowledge sharing, knowledge use, and knowledge creation.

This paper is organized in four sections. The first one discusses the importance of KM to KDD and vice-versa. The second one outlines the organizational context where the case study was developed. The third one depicts the applied methodology [13] based on the trade-off between the KM Model of Stollenwerk [14] and the CRISP-DM methodology. In addition, the case study is also described in this section. Finally, the fourth section presents some findings of this research and points out some future works.

2. KDD AND KM

According to Peter Drucker [3] the most important challenge posed to the knowledge society is the development of systematic practices to manage the self-transformation that is based on organizational learning. In the same direction Senge [12] argues that the manager should follow five disciplines to develop an organization that learns: systematic reasoning, personal control, mental models, sharing views, and team learning. Moreover, KM requires an intense human interaction to generate, storage and share organizational knowledge.

KDD can be seen as any other process that needs to access the burden of organizational knowledge in order to have each task performed efficiently. Davenport and Prusak [2] emphasize the fact that KDD technology is part of KM, working as a lever to it. This technology plays an important role as it brings a methodology for the expert to perform the extraction of knowledge from the available databases.

On the KM side, Stollenwerk [14] discusses the existence of many models to describe KM. The author stressed these models looking for a synthesized alternative and, in all of them, identified the processes: identification/conceptualization, acquisition, selection and validation, organization and storage, sharing and transference, application and use, creation, products and services engineering based on new knowledge, and evaluation of the benefits and the value of the knowledge. On the other hand, Sveiby [15] classify the KM approaches as regards information management and organizational learning and abilities management. Based on these ideas, Stollenwerk coined a generic model, which focusses on organizational learning and abilities management. Our work is based on this model, which we extended [13] to show the importance of KDD to KM.

3. ORGANIZATIONAL CONTEXT

The Electoral Court of the Federal District (ECFD) belongs to the Brazilian Justice and is in charge of the election process in the Federal District, ensuring law enforcement. The Court has performed important services to the community and the parties.

In the KM context, the ECFD is working to map its human potential to generate knowledge by using the internal abilities and a set of databases. Also, a mapping of the external requirements has been carried out that cause an important impact in defining changes in the internal process. As shown in Figure 2, KDD can be naturally integrated to the KM model of ECFD , that enables the activation of the mental processes [9] that, in turn, lead to pro-active interventions from the Court board.

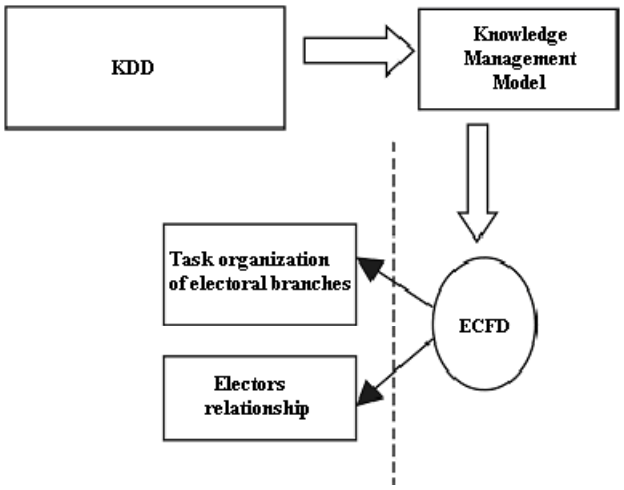

Figure 2 - Relation of KDD and KM

It is important to emphasize how crucial the involvement and the support from the board to the knowledge creation process is. According to Terra [16], the definition of knowledge areas to be explored by the organization and the statement of general views that can conduct innovative projects are part of the fundamental role of the board.

4. CASE STUDY

The case study comprises the application of our methodology, distilled from an analysis on the relations of KDD and KM. This analysis, carried out by Sousa [13], produced an integrated model (see Table 1) by comparing the steps from KDD with the generic KM model of Stollenwerk [14].

Table 1 – Integration of KM and KDD

#	Integrated model	Generic models	
		KM Model [14]	CRISP-DM
1	Problem and context definition	Identification of abilities	Business understanding
2	Capturing the required inputs to the problem solution	Knowledge capture	Data understanding
		Selection and validation	
3	Processing the inputs to the problem solution		Data preparation
4	Defining techniques and algorithms for knowledge extraction		Modeling
5	Expert assessment and interpretation of data mining results		Assessment
6	Structuring the acquired knowledge	Organization and storage	
7	Publishing results	Sharing	Deployment
8	Effective use of the acquired knowledge	Application	
9	Global assessment	Knowledge creation	