Concept learning – investigating the possibilities for a human-machine dialogue

TECHNICAL REPORT

Gabriela Pavel
January 2009

Part of the probation report (11th December 2008):
"Intelligent support for learning concepts from examples"
Supervisors:
Dr. Zdenek Zdrahal
Dr. Paul Mulholland

Table of contents

Introduction

Motto: "*Education* is the kindling of a *flame*, not the filling of a *vessel*."

(Socrates)

Everyone around us learns. As a flower needs to adapt to the light, an animal to its environment, we people need to adjust to our complex social or personal conditions of life. Because of that we definitely learn during our entirely life. So, we could say that is better to start learning anytime. We could say that learning process is one of the most important parts of our existence. For examples teenagers are more likely to be receptive to the learning environment than older people. Then, what needs to be done for a more productive approach of society development, if not supporting young students and children to improve the learning methods? If we accept the learning process as defined by the acquisition of concepts or of competences, then we will have to reformulate the general question: how to support youth in learning concepts?

Not a long time ago, education was a filling process. Students were supposed to acquire knowledge and learn it by heart. Well, education is more than this. Furthermore, how about the ability to recognize and explain what is learned? That means to understand new learned concepts and be able to use them in contexts. A good teacher recognizes the potential and supports the students to develop and learn new competences and new concepts. Nowadays, technology is used to help people learn faster by providing them all the necessary support for the communication in the educational process: distant resources access, enabling access to other human resources (oral dialogue is a major component of human learning process).

What is above represents the premises of our research. Before this, we need to mention that we will focus on learning from examples. Why? Because this is the way young people develop: by imitating the adults. That is why it's very important to know what examples to offer, how to present them and how to support the learning from examples. Children are required to learn concepts; concepts are met in everyday life and the curricula ask them to learn as well. Sometimes concepts are quite complex; can we take a simple case like the green colour and claim that we know everything about it? Maybe, the RGB numerical representation of the colour is accessible (#00FF00 is the RGB representation of green). But, "green" can also be defined as an attribute for life, or metaphorically as hope. What about poetry? Can we identify precisely the mind role during this process? Definitely there are a lot of questions to be answered. As we see, it is not so easy to define a concept unless we are at least familiar with the context. So, if we intend to use technology in supporting the learning of concepts, we wonder: why not using AI to support the learning from examples?

After this small journey through some general questions concerning the learning process, it's time to introduce the aim of this report: to investigate the potential of machine learning algorithms in supporting the human learning of concepts starting from a particular set of examples: images. But, first let us introduce our problem.

1. Problem Specification. Research goals

In this chapter we will present first the problem and the questions derived from it. The research questions and the plan for answering them will be presented in another paper. So, let's see now our research problem and the context of the investigations presented in this report.

1. Problem specification

Our research is focused on computer-based support for human concept learning. Let us define our problem by answering three questions:

a) To **whom** do we address our work?

We want to help the human learner explain the description of an acquired concept. A particular case of people is formed by youth with ages ranged between 11 and 14 (form the Key Stage 3 group of study).

b) **What** exactly do we want to do?

The learner intends to learn concepts using set of positive and negative examples, and in particular, images. We want to support the learner in articulating the concept from these examples. Using the computer for this purpose is a challenge, because the formal description of a concept is usually a set of (attribute-value) pairs, and by consequence a poor description with very limited flexibility (algorithms accept however missing or unknown values for attributes in the concept description). We want the computer to support the articulation of concepts which are not easily expressed, for instance social concepts, such as "democracy", "right to vote". People are unable to define them using a set of attributes like "gender", "place", "race" etc. because they are not able to articulate such concepts (counter-examples can be easily found, and by consequent the description will be incomplete and it will be likely to be incorrect; we will see this later in this paper). Even "simpler" concepts like "object being round" cannot be described only by the perfection of its geometrical shape; we will see later in this paper (when talking about typicality) that some objects are more "round" than others; for instance, we can consider a "ball" to be more round then the "earth", which is more round than an "apple".

c) **How** do we intend to support the learning process?

We assume that the learner can be supported by ML (machine learning) techniques in the learning of concepts. ML techniques are used in their traditional way for learning descriptions of simple concepts. As we will see in the next chapter, ML has potential for our problem because of its ability to deal with typicality (for instance, conceptual clustering treats the border cases of more complex-defined concepts).

2. Context of study

This work is part of the SILVER project, which is interested in supporting the learning from examples. SILVER has access to the digital collection of images provided by the Bridgeman Archive from London and has as purpose to build technologies which will support both pupils from key stage 3 (with ages between 11-14 years) and teachers in the learning of complex concepts. The images are supposed to be used as they are and with metadata (attribute-value pairs) extracted from the set of already existing annotations about their content, period of time and author. The project is focussing on citizenship topics, like democracy and voting.

Citizenship involves the communication with the purpose of understanding the role in the society (Higgins and Packard, 2004) by talking to or listening to others. Higgins and Packard (2004), talking about the teaching of the citizenship topic in schools underlines the mechanism of choice (the second unit of the curricula being focused on choices) and the use of picture/video to learn. The first unit of the curricula is focused on communication skills in the form of illustration of ideas and resolution of conflicts inside the classroom.

Being part of this educational project, our research will focus, as mentioned before, on how ML will support the human in learning from examples. That is why, our goal is to interconnect two types of learning: human and machine learning and how the ML techniques can be used to support the human learn complex-defined concepts; therefore, we need to analyse both processes of learning. First, we will analyze the human process of learning, more precisely the concept acquisition; then, we will focus on how people articulate knowledge in the reasoning process. We will present afterwards an overview of AI techniques that can be used in eLearning and the limitations of existing ML techniques. We will select three ML techniques that we consider to have a potential for supporting the human learning process. We will present possible extensions of these algorithms to support the human learning and how we can merge them to support the visualization of the reasoning process. Then, we will present our proposal: a mutual feedback between human and machine with the purpose of helping the human learn; we call this approach the *Human-Machine Concept Dance* (HMCD); we will present the two types of feedback involved in this approach (from human and machine). Finally we will present the limitations and the permissions of this approach, as they are revealed in this early stage, as well as the future plan for investigating deeper the possibilities of this approach.

3. Research goals

Given the intrinsic complexity of concepts, we have two possible dimensions of our research problem related to this aspect:

1. how much of a concept meaning can be formalized (described as (attribute-value) pairs)? We are interested to see during our research how much of the concept description, offered by the learner, can be represented in the computer.
2. how to control the evolution of the description using ML techniques? We assume the descriptive representation of a concept will change because of its complexity.

   Our goal relates to how much can the ML influence the learning process and how much the user can influence the ML-based learning tool (because the ML techniques will use formal representations, as they are given by the human learner, and ML should offer answers that can be understood by the learner).

4. Structuring the Literature Review

In the next chapter we will provide an overview of the most relevant references for our work. Here we must address the following issues: *how* ***machine*** *learns concepts*, *how* ***people*** *learn concepts*. We start with a presentation of different views on concept definition and representation. Afterwards, we try to answer our main question: *how to support the learner in learning concepts?* We know that the basic representation of concepts (attribute-value pairs) suggested us the use of machine learning techniques.

ML algorithms use the basic conceptual structure and identify concepts from the set of examples provided to the system. At this point of the review we can understand that sometimes concepts cannot be formalized.

Rosch (1999) introduces the notion of “natural concept” which is of great use for us; Rosch associates the complex representation of concepts with *the graded-structure* (or *typicality)* that allows objects to be grouped around the prototype (i.e. the best instance of the class); the grouping process regards the semantic similarity of objects to the prototype. Still*, how to formalize what it cannot be entirely expressed*? After this finding, we may seem to change the strategy: no use of machine learning anymore?! In addition to this, we see the two dimensions of knowledge: tacit and explicit (which explained the human inability to articulate completely the description of a concept). However, machine learning techniques are also used to deal with typicality and that’s how we could finally define our proposal. Moreover, tacit is shared in a collaborative environment, so, the classroom will support the learning with its both dimensions: implicit and explicit.

We investigate also previous approaches of AI techniques in education, to see how to adapt AI to the human learner in order to improve the human-machine collaboration.

Concluding, there is a niche for us and a chance to use ML and “natural concepts”; the remaining problem is *how to support the human learn concepts from images using ML?* This is what we hope to answer during this PhD. Now, it’s time to start presenting our review.

2. Literature review

In the present report we are interested to answer to the challenge of using AI to support youth learn concepts from images. Because we accept learning as a process by which people acquire concepts, we need to define then; in this chapter we make an overview of different approaches on concept definition and representation. We will see then that there are “natural concepts” (Rosch, 1999), which are difficult to define. We will focus on three ML techniques that we consider suitable to support the learning of concepts from examples: vector spaces (Mitchell, 1978), decision trees (Quinlan, 1993) and generalized logic diagrams (Stepp and Michalski, 1986); we will investigate also different approaches of ML techniques that try to deal with the more complex represented concepts. Afterwards, we present some related work concerning the use of AI in education. Then we present a logical framework for the learning process.

1. Concepts

Concepts are, according to (Klausmeier and Allen, 1978) “mental constructs of the individual and the socially accepted meaning of one or more words that represent that particular concept.”; by this, the authors are defining a concept as an agreed understanding of a collection of individual objects. According to these authors concepts are the property of a class of objects and therefore concepts represent a way to categorize things around a common meaning/idea (Braisby and Gellatly, 2005). Braisby and Gellatly mention the importance of concepts in problem solving, for two reasons: a) concepts help retrieving information and farther use this information, and b) they build blocks of knowledge and by thus are used in learning new concepts.

Rosch (1999) presents and analyses several different views on concepts: 1) the classical view; 2) the prototype view; 3) the cognitivist view; 4) the new approach. We are interested in these approaches, as they can help understand the complexity of defining a concept. Because of this, we will characterize now some parts of Rosch work. We will start with the classical view which introduces the basic description of concepts.

Simple view on concepts

According to the classical view on concepts, as it is presented in (Rosch, 1999), concepts are built to answer questions. In this view, concepts are represented as logical rules applied to attribute-value pairs.

> E.g.: a “red triangle” is characterized by the attribute “colour” with the value “red” and the attribute “shape” with the value “triangle” (Bourne, 1970).

This point of view is shared by Bourne (1970) who defines a concept with a rule and some attributes. If C is a concept (class of objects), then it can be represented as a rule applied to a set of (attribute-value) pairs: $C=R(a_1,\ldots,a_n)$, where $a_1,\ldots,a_n$ is a set of (attribute-value) pairs (binary attributes, so they can have only two possible values: true or false) and R is a rule. In Bourne’s perspective, a rule is a logical formula of the first order logic that may include conjunction, disjunction, implication, double implication (known as equivalence) and negation. Attribute is the representation of a feature. Given this representation of concepts Bourne identifies two problems in the concept learning process: 1) identify the rule knowing the attributes (know how to combine pairs of attributes and values in order to form rules/relation) and 2) identify the attributes knowing the rule. Bourne gives two conclusions useful for our learning problem: first he says that few examples are necessary to make the human learn a concept (the subject is able to deal with unseen examples, using the provided examples). Second, Bourne concludes that concept learning allows the transfer of conceptual properties between different levels in the conceptual space (a conceptual space is formed by rules, exemplified by concepts or classes; in a conceptual space attributes are properties that define those objects that belong to a concept).

In the classical view, (Rosch, 1999) mentions a set of properties that are used in defining concepts in a simplistic way:

1) concepts have clear boundaries (the relations between an attributes and its values are well defined. By consequent, two examples characterized by the same (attribute-value) pairs define to the same concept);

2) have necessary and sufficient common attributes to guarantee the membership in the category;

3) are equally good: concepts have all the necessary common features or not.

In contrast to these "simple" concepts, the psychologist introduces the "natural concepts" (Rosch, 1999). Unlike the simple concepts, natural concepts are complex; they might be expressed using a simple set of (attribute-value) pairs, but it is difficult to know which set to select from the beginning; by consequent, complex concepts cannot have clear boundaries. In addition to this, concepts contradict the other two properties mentioned above for the definition of simple concepts: there are no necessary-sufficient conditions to define concepts, because we think that an attribute can appear in different concepts and still be a characteristic of the defined concept. The membership cannot be guaranteed by the presence of a certain set of attributes; attributes should be considered after the categorization, they being analysed only the context within they appear (Rosch, 1978). So, as we see, this traditional view on concepts is of no use for our research problem, when we are interested to support the learning of natural concepts (concepts that we meet in everyday's life).

Another view on concepts presented in (Rosch, 1999) is the *prototype view* or *graded structure view*; this perspective considers that "all categories show gradient of membership" and associates ranks of membership to the objects (Rosch, 1978). In this perspective, a prototype is characterized by the properties that define the object with the highest degree of membership to the class. The "better examples" are used to accelerate the speed of the processing and the learning of a concept (Rosch, 1999). The positive concepts are given in the context of knowing in advance the expected category and help the learner make associations and inferences. This perspective is also mentioned as "typicality" and this is the kind of representation that we will use to define concepts (natural concepts).

Smith and Medin (1981) consider a similar point of view with the representation of concepts by rules and (attribute-value) pairs: dimensions and features. For these authors, the dimensions are what we have already seen: (attribute-value) pairs. In addition to this, they used a probabilistic approach to identify which dimension to use in the definition of a concept, according to the degree on which a dimension contributes to the class membership. In their probabilistic experiments, they identify the presence of:

- unclear cases (whose dimensions are close to the threshold and by consequent they cannot say clearly to which class the described object belongs);

- nested cases (considered like this because they are similar to their immediate super-ordinate; they give the example: *robin* is a *bird*. There are exceptions from this rule; for instance a concept can be similar to a more distant concept in the hierarchy than its direct parent: *chicken* is closer to *animal*, then to *bird*).

The features-based descriptions of concepts can specify properties, according to (Smith and Medin, 1981). This view is close to the graded structure perspective on concepts and we can affirm that it is also related to the representation of complex concepts.

The third view on concepts is given by the theory of cognitivism (Rosch, 1999), which we won't develop here.

The most interesting view is the new perspective on concepts, presented in (Rosch, 1999) in terms of a set of properties of concepts:

- concepts are "bridges" between world and mind: concepts are not the representation of objects in our mind, but they take part to the relation between objects in the world and objects in the mind;
- concepts appear in concrete situations;
- concepts never occur in isolation;
- concepts don't identify situations, but they take part into situations.

This new view on concepts is applicable to natural concepts, which don't offer definitions of the world, but help us understand it.

There is a way in which learning concepts affects *memory* organization. Concepts from the same domain are stored together in the memory. The subject has the tendency to recall either the whole "chunk" of elements or none (Anderson, 2000). Given this associated representation inside memory, Anderson recommends the use of associated lists of visual objects (subjects can associate either objects that interact, or objects that cannot – in the last case, the subject can select objects that dissociate with others in the list). Memory is important also for its use in acquisition processes (the retrieval of information) as mentioned in (Reber, 1993), in the form of discovering rules (patterns). For instance, in learning algebra equations, we can discover a pattern like the following: if on the left side of the "=" sign we have two numbers/equations connected by the "+" operator", on the right side of the equal the subject will complete the sum of the two numbers separated by the addition operator (Zhu, Zhu, Lee and Simon, 2003).

Concepts are *organized hierarchically* in semantic networks in the memory according to (Borger and Seaborne, 1982) and (Murphy, 2002) from where they are recalled in two steps: retrieval and decision. Because the memory is limited it is not possible to store there everything related to a hierarchy; by consequent, the *cognitive economy* function of the memory allow to the subject to store the differences and rebuild the whole hierarchy when needed. Concepts are grouped hierarchically within a taxonomy (Rosch, 1978). A taxonomy "is a system by which categories are related to one another by means of class inclusion." (Rosch, 1978). We think that the hierarchical descriptions are important in learning concept, because they permit the use of connections to make inductive inferences and categorization judgements, which can be useful in our project, as part of the learning process.

Feedback

"Action without feedback is completely unproductive for a learner" (Laurillard, 2002). From this perspective, learning is to adjust the action to the feedback. The importance of feedback is in the way it is received and used. Feedback can be:

- intrinsic: the natural consequence of an action; this seems similar to the process of transferring tacit knowledge into tacit knowledge;

- extrinsic: an external comment (e.g. right/wrong) to describe actions, external to the context of the action (e.g. the use of an alternative description known by the student).

Feedback should be relate to the goal, in the cycle goal-action-feedback (relate to the structure of the whole).

Lieberman (1987) analyses the learning from examples in which the visual representation of the effects of learning can help understanding the mistakes. It is important to know which attributes are important for an example and that the learning sequence should start with a simple example and add more complex examples.

In education, the Information Communication Technologies (ICT) provides the use of multimedia in the learning process (Laurillard, 2002). Students are taught to use in addition to text, video and audio materials, and also pictures, in order to acquire knowledge. This sort of material provides interactivity, but students are not able to communicate what they learn. They can be evaluated by tests (online tests are very frequent) if evaluation is some sort of communicating what they have learned. But we are interested to see what happens during the learning process. That's why our interest in this paper is on the investigation of different ML techniques and how they can support educational strategies (we are interested in learning from examples, and in particular, from images). First we will focus on the tacit-explicit dimensions of knowledge as they are related to the process of articulating knowledge. The acquisition of a concept can be proved either by doing (for instance, selecting some images), either by communicating it verbally. In the first case, this is explained by the tacit dimension of the learning process (the skills proved in doing). In the second case, this is made by externalizing the knowledge (the transfer from tacit, what is known, to explicit, what is explained).

Then we focus on the process of learning from the two perspectives: the machine and the human. We will make an overview on some AI approaches used in education. Finally, we will focus on our research proposal and how we think to evaluate it.

2. Dimensions of knowledge: tacit versus explicit in the human learning process

Knowledge represents more than its explicit aspect. There is a possibility that our knowledge is not richer then what we have expressed. "We can know more than we can tell" (Polanyi, 1967:4); what we can't tell is the tacit aspect of knowledge. Polanyi introduces the notion of tacit knowledge: "certain knowledge that he [a person] cannot tell" (ibid, p. 8). The basic structure of tacit knowing involves two components: the first component is condition; and the second is an action as a response to the condition. These two components are combined in order to provide a "meaning" to the action. Polanyi introduces too the tacit knowing as process and explicit knowledge as state.

We consider both dimensions of knowledge (tacit and explicit) very important in knowledge acquisition. First, we will focus on the process of creating knowledge in an educational environment. Then we will reveal the relation between knowledge creation and the learning process. Finally we will discuss the advantages of using the dimensions of knowledge in an educational environment.

The knowledge creation is "a process of making tacit knowledge explicit" (Nonaka, 1994). The steps of knowledge creation are:

1. metaphors = use of imagination and symbols for analysis and generalisation; metaphorical images have multiple meanings and are often contradictory (connect things that seem unrelated).

2. analogy = reconcile contradictions and make distinctions between the new image and the existing information;

3. create an actual model = solve contradictions and transfer concepts.

From the point of view of expressing it, knowledge has two dimensions: tacit and explicit. (Nonaka, 1994) defines *tacit* as "not easily expressible" and deep in context (for instance it depends on the technology used), "know-how". On the other hand *explicit* knowledge is "formal and systematic".

The dimensions of knowledge play an essential role in the transfer of knowledge in a social environment. School is more then a social environment, because it can influence the development of future generations. That's why, we will be interested in this report to evaluate the transfer of knowledge in students and how the computer can be used to stimulate this transfer and help the students learn quicker.

The description of the knowledge transfer the transfer of knowledge, also called "the spiral of knowledge", is presented in table 1 (Nonaka, 1991). The square allows four possible transitions.

		To	
		Tacit	Explicit
From	Tacit	Socialization	Externalization (articulation)
	Explicit	Internalization	Combination

Table 1. The spiral of transferring knowledge

Below we will explain the four pattern of knowledge transfer from the perspective of the teacher-student, student-student dialogue:

1. *Tacit to Tacit*: students learn by observation and practice; students see how the teachers or their peers solve an exercise and learn how to do it.

2. *Explicit to Explicit*: synthesize information (students make a report after a new topic is introduced using previous explicit knowledge).

3. *Tacit to Explicit*: articulate own knowledge; the process is called articulation and consists of the use of figurative language, metaphors: "express the inexpressible" (Nonaka, 1994). Students explain (are able to verbalize) their intuitive knowledge. The verbalization process is quite complex and we will underline it in the next section.

4. *Explicit to Tacit* (or the process of internationalization): sharing knowledge so that another person can internalize it.

As we will see immediately, there are very different opinions on the process of transferring tacit (T) knowledge into explicit (E). There are different views on the possibility of verbalizing tacit knowledge. According to Berry and Broadbent (1984) tacit is supposed not to be verbalized; subjects can't explain are unaware of the process of learning. But, in this case, we may wonder how can we talk about verbalization (or articulation – as Nonaka mentions it) if tacit has no chance to be expressed?! Another point of view is too flexible: Nonaka (1994) assumes that tacit "express the inexpressible" (T becomes E) (should we think that any tacit knowledge can be articulated, or there is still something that we cannot explain (Polanyi, 1967) in the tacit dimension of knowledge?!). Polanyi, on the other hand, thinks that tacit can be communicated by adequate means. Sternberg and Horvarth (1999) believe that tacit is "rarely openly expressed or stated" with which we agree because tacit itself cannot be expressed completely, but it has a component that can become verbalized in order to make the transfer of knowledge possible. "Openly expressed" knowledge will be in our opinion the already articulated knowledge, which is by definition explicit. We think that tacit knowledge is "unconscious acquisition of complex knowledge" as Reber (1993) says, but that it has an expressed part that makes the verbalization available. Certain part of the tacit can be expressed.

We develop explicit knowledge which will create tacit knowledge (for instance, by articulating acquired knowledge people develop both explicit knowledge and the skill, which has a tacit part, to recognize similar pieces of knowledge).

Trying to synthesize the ideas on tacit knowledge, we can represent the degree of expressing/articulating tacit knowledge (figure 1):

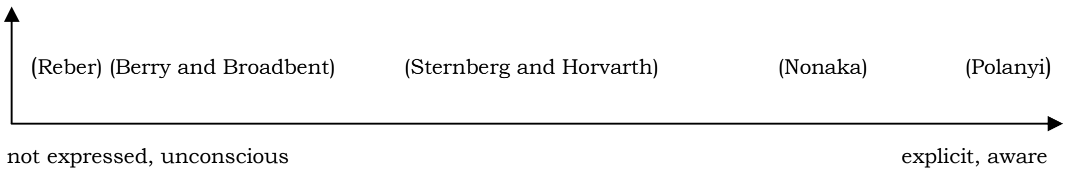

Fig. 1. The verbalization of knowledge, different views

The tacit-explicit dimensions of knowledge play a role in the learning process, as we can see in the next section.

Implicit and Explicit Learning

Implicit learning is a natural and unconscious acquisition of knowledge about the structure of the environment, while *explicit learning* is the conscious operation of making and testing hypothesis in search for a structure (Reber, 1993). The study of the second language acquisition in this context requires the implicit and explicit dimensions of learning as the incidental and intentional dimension of vocabulary acquisition.

We can extend it also to complex concept acquisition. Implicit learning is: “a) unintentional and thus incidental […] and b) involves induction without awareness" (Reber, 1993). Conscious, on the other hand, is defined as intentional, the product of attention, aware, instructed, controlled (ibid.).

Both implicit and explicit learning are involved in the incidental vocabulary acquisition (Rieder, 2003). Extending the idea, we can argue that both incidental and explicit knowledge are involved the process of learning complex-defined concepts: the implicit learning is the learning without being aware of a concept. The explicit learning is involved when the student is required to describe it in terms of (attribute-value) pairs.

All this is related to the articulation (Nonaka, 1994) and we wonder how this process develops in order to model it in the computer and help the student verbalize the internal knowledge quicker. But, before trying to answer to this question, we may need to analyse the process of learning.

Dimensions of learning

Learning is defined as an acquisition process and all steps in the knowledge spiral are involved here in one way or another.

First we see how psychology defines learning. For constructivists, learning is a "permanent change in mental associations" (Piaget, cited in Prichard, 2005). If we put an equality sign between the two definitions, we can conclude that the acquisition of a concept means its integration in a mental scheme (Scheme theory). In Piaget's opinion learning is a process in 3 steps:

1) assimilation = accept new information as it is;

2) accommodation = integrate the new information in the existing information;

3) equilibration = the process of solving conflicts generated by the accommodation.

Seen as a process of having, integrating and solving information, the learning process is similar to the 3 steps of knowledge creation proposed in (Nonaka, 1994): metaphor, analogy, model (or the adaptation).

The creation of knowledge as a transformation from tacit (symbolic images) to explicit knowledge is similar to the process of learning seen from the constructivist perspective. Given this analogy we can conclude that the *knowledge creation is a process of learning*. The similarity between the knowledge creation and the learning process is represented in the table below (table 2):

Knowledge creation (Nonaka)	Learning (constructivism: Piaget)
METAPHOR	ASSIMILATION
ANALOGY	ACCOMODATION
MODEL	EQUILIBRATION

Table 2. The similarity between knowledge creation and learning

The dimension of tacit in the learning process

Sternberg and Horvarth (1999) analyse the tacit knowledge involved in learning. According to these authors there is a tacit dimension in:

- the nature of the discipline (a domain knowledge); this type of tacit is the tacit defined as “deep in context” (agreed also by Nonaka, 1994; Polanyi, 1967);

- the nature of learning (where the teacher needs to know that learning is an active process, and students have a limited memory capacity);

- pedagogical knowledge (the knowledge about how to teach: the "know-how" of the teaching activity);

- curriculum design (the "what to" say part of the teaching activity).

"If tacit knowledge needed, then it must be generated and made sharable. Or both." (Cook and Brown, 1999). That's why in an interactive environment is useful to use the dimensions of learning and the classroom strategies to generate new knowledge. Cook and Brown call the formation of new knowledge "the generative dance". This knowledge formation involves both the knowledge as action and the knowledge as explicit acquired information.

In the individual learning the explicit knowledge will be reflected in the acquisition of concepts which introduces the definition of learning as "acquisition of explicit knowledge" (Cohen and Feigenbaum, 1982). The tacit component of learning is reflected in the learning seen as "skill acquisition" (Cohen and Feigenbaum, 1982).

In the collaborative aspect of learning, the tacit-explicit dimensions of knowledge work in a similar way: the explicit knowledge helps sharing explicit information and by thus the formation of stories. The tacit help the group share an interpretation, a meaning and build the so called "genres" (Cook and Brown, 1999). Because of this property of making explicit knowledge available among its members we will focus on the collaborative learning.

In the transfer of knowledge, the knowledge is identified simultaneously as action and possession (Cook and Brown, 1999)

- *knowledge* (enables possession of explicit knowledge);

- *knowing* (requires present activity. “Knowing” is the "knowledge in action"). Polanyi writes about "tacit knowing", which will give to the tacit the possibility of evolution and explaining itself. Cook and Brown are against this association between tacit and knowing because in their opinion tacit doesn't require "present activity" (it contains only skills, a meaning possessed by the individual or the members of the group). We share this last opinion and consider that knowing can use tacit only as a tool (the use of skills in action for generating new knowledge and action).

Below we represent the knowledge that can generate new knowledge if combined with knowing, as presented in (Cook and Brown, 1999):

	Individual	Group
Explicit	CONCEPTS	STORIES
Tacit	SKILLS	GENRES

Human Concept Learning

Some researchers mention two *types of learning*: *learning without awareness* and *incidental learning*. In the process of problem-solving, people seem not to remember what they knew or what they used; subjects show in an implicit/indirect way to have “a memory residue of an earlier experience”, an *implicit* form of *memory* (Reber, 1993). On the contrary, the *explicit memory* contains instructions which can improve the learning (even if the instructions are received at the beginning of a learning experiment, of after a set of examples is presented, ibid.). We may ask therefore how to use the tacit learning in a collaborative environment in order to support the concept acquisition.

Besides the understanding of learning as a process of acquisition of explicit knowledge, there is another important aspect of the human learning: the learning as skill acquisition (Zhu, Zhu, Lee and Simon, 2003). This sort of learning is related to tacit knowledge, which is, in this situation, represented in the form of productions that student develop after seeing a set of examples. These productions are not necessarily expressed verbally, but they are activated when the required conditions for the activation of productions are accomplished (ibid.).

Before continuing with ML techniques, we may find it useful to know how to design a human learning scenario, which involves the building of a “***learning design***”. Weller (2007) defines the *learning design* as a framework for the description of methods used to interchange teaching strategies and learning objectives. Using Weller’s perspective, a frame for a learning strategy useful for our research consists of four steps: we define a problem, we build the scenario (a model, as it is presented in the next chapter), we define a descriptive vocabulary and we use a ML-incorporated new tool to support the students learn from images. So, we will have to design a system that uses the investigated ML techniques for supporting the human in learning concepts from examples (we will see this in another chapter).

Now it's time to see how ML techniques use concepts.

3. Machine learning techniques

As we have seen previously, we are interested in the externalization process as a way of providing a clear description of the concept (especially for a complex-defined concept, such as "democracy"). That's why, we analysed first concepts and the dimensions of knowledge. Now, we all focus on the learning from the machine point of view.

Learning is defined in (Cohen and Feigenbaum, 1982) (we have already seen two definitions accepted also in ML as in human learning):

1. Learning is "any process by which a system improves its performance" (Herbert Simon).
2. Learning is "acquisition of explicit knowledge"
3. Learning is "skill acquisition".
4. Learning is theory formation, hypothesis formation and inductive inference.

From the point of view of the learning environment, in ML, there are more types of learning (ibid.):

1) rote learning

2) learning by instruction (by being told) which can be expressed in two ways:

   a) learning from examples (learning by induction)

   b) learning by observation and discovery

Learning by instruction involves five steps: request, interpret, make operational, integrate, evaluate. The last step is used as a feedback to the request and the learning process can be run again (Cohen and Feigenbaum, 1982).

3)learning by deduction

4)learning by analogy

As we have seen in the previous section, human represents concepts in a complex, flexible way. ML requires explicit description of instances of a *target* concept (concept to be learned). We will present three ML techniques that we intend to use partially because of their visual representation and second for their potential to express a flexible concept description. We think that the visual representation of the learning process can help understanding the reasoning process of the learner. The techniques that we investigated here are version space, decision trees, generalized logic diagrams.

Version spaces are used in state generation (version spaces are used in stories-based learning in physics, as we will see later). Decision tree can be used for improving human understanding and readability (Shapiro and Niblett, 1982). AQ21 is used for its readability (develops mathematical thinking; that's why its potential for education should be investigated).

a) Version Spaces

A conceptual space contains concepts that classify correctly the training instances. The concepts are called hypothesis as they are conjunction of constraints and they cover those examples that respect those constraints.

A hypothesis space is ordered from general to specific. The learning of a concept is a search in the hypothesis space.

The *candidate elimination algorithm* (Mitchell, 1978), used to acquire a concept from positive/negative examples in a version space, allows the use of two operations: specialization (S) and generalization (G), which add new hypothesis to the maximal boundary sets: S and G.

Both sets are computed for a positive or a negative example in the following way:

1. for a positive example: G is enriched with those hypotheses that cover the new instance of the target concept; in the same time, S is updated to cover new concept;

2. for a negative example: S remains the same, while G eliminated those hypotheses that cover the negative example.

Hypotheses change in a version space in order to cover new instances of a target concept. If we consider a hypothesis to be the representation of our *belief* on what possible concepts the instances can lead to, then we ca associate the hypothesis formation with a "*belief revision*" (Hansen, Pigozzi and Torre, 2007).

But, first let's see how deontic logic works:

> For instance, students must express verbally an acquired concept. This sort of requirement is an obligation. If a student is allowed to see a set of examples before being able to verbalize the target concept, then this is a case of permission; this is a case of learning by examples, which is different from learning by doing by the fact that the examples can be sorted by the teacher (Zhu, Zhu, Lee and Simon, 2003), while in the second case of learning, the student himself/herself identifies what type the example is: positive/negative, by working with them; so, we think that learning by doing cannot precede learning from examples. In this case permission was just helping the student to complete the obligation: the student learns by examples and by doing and then is able to present what he/she have learned. However, permission can act like an exception from an obligation: a student is allowed just to recognize images representing the target concept, not necessary to express it verbally (that is the important is just to demonstrate the acquisition of a skill, which is a tacit form of knowledge, and by thus, not necessarily verbalized).

In deontic logic[1], more exactly in normative systems, a *belief revision* formalizes the changing of propositions according to the new information that can cause inconsistencies in the existing beliefs (what we think about something can chance because of additional information). Belief changing comprises three types of operations:

1) expansion: we can assimilate it with the specialization operation in a version space (adding new specialized attributes);

2) revision: modify attributes, values or both (still under question, as it belongs to the possible extension of the version space);

3) contraction: generalization (less specialized attributes).

The three operations used for the process of reviewing a belief system is very similar to the way in which a hypothesis updates in a version space (updating through three operations:

[1] A **deontic logic** is a system which capture features of the concepts of: obligation, permission and prohibition.

generalization/specialization/change; the last operation concerns the resolution of contradictory cases by modifying/enriching the descriptive vocabulary). Because we have mentioned the contradictory situations we need to explain them.

Contradiction identification

In an ideal version space, every positive example is an instance of the target concept. A positive example is described by a set of features and some of these characterize also the target concept; these last features are relevant features, because they help identifying the target concept. On the contrary, a negative example has no relevant feature. Learning by presenting counter-example is useful because it reflects the change of understanding a concept (Kang, Scharmann and Noh, 2004); these authors are interested in teaching science concepts, such as "density", using *conceptual conflicts*, which bring equilibration in the learning process. Students are also able to discover inconsistencies in their definition and learn about the complexity of concepts (ibid.).

The contradictory situations from a version space are similar to the contradictory situations inside a space of o*bject*s instead of *concepts*. Sussmann (1975) identifies conflicts *inside a space of objects*; he makes investigations for a PUT-ON strategy and identifies the presence of a *GHOST* (= "*object that cannot be moved*" – Sussmann, 1975:55). In order to avoid such situations (*contradictions*), he tries to *solve first all conflicts* for the next set of moves and *then make the moves*. Sussmann defines also the *generalization* with two components: the result of learning from experience and the ability to apply substitution for an entire procedure (ibid, p. 84).

Coming back to version space, we have to mention that a poor descriptive language may permit to a positive example to lead to a concept other than the target concept. If the positive example contradicts a negative example, then a contradiction is likely to occur. In the case of the contradiction, S is null. If S belongs to the E (space of eliminated hypothesis), then we deal again with a contradiction. A contradictory example requires the elimination from H (the hypothesis space) of a hypothesis which is already in E. Therefore for a positive example, build first the E, and then the H.

The ideal version space works with conjunctive concepts. A conjunctive concept can be represented as a conjunctive normal form. Contrary to this, the disjunctive concepts can be represented as a disjunctive normal form; the candidate elimination algorithm cannot be used to learn disjunctive concepts. The disjunctions can be solved in multiple version spaces. Previous approaches defined operations on version spaces that allow the addition/removal of a version space in order to represent the disjunctive concepts.

Disjunctive concepts in Version Spaces

Intermediate concepts solves, according to (Hansen, Pigozzi and Torre, 2007), juridical conflicts (for instance the presence in a set of admitted terms of both "dog" and "non dog" is a contradiction, while the replacement of the second "dog" with "guide dog" solves it). We can understand this solution as an act by which we replace a concept with a subordinate concept, more specific (the case of "dog" and "guide dog"). Should this strategy work also in version spaces? We hope it can help to understand the building of intermediate concepts. Intermediate concepts applied to laws "link legal terms to words describing natural facts" (ibid, p.16). In a similar way, we consider that intermediate concepts in a version space link the generalization/specialization operations to words (attributes or values) describing facts that contributed to the presence of contradictions (certain values for a given attribute). But, while in a normative system, intermediate concepts involve the presence of the "Ownership(x) term" to reduce the number of implications of the consequences of a fact, in a version space, intermediate concepts could add more hypothesis in S and maybe in G as well, process which is still unclear; we aim to investigate this issue in our research work.

Intermediate concepts are useful for introducing additional nodes in a version trees. Fu and Buchanan (1985) introduce two types of intermediate nodes:

a) containing values already existing in the initial descriptive vocabulary (in which case a redundancy check is required);

b) containing values not present in the vocabulary (generic names for clusters of nodes, where nodes are grouped according to their similarity).

This idea of this strategy is very useful for our project, but in our case we intend to use values (not generic names) not present in the vocabulary. We will see how we intend to find the new values in the proposal section.

Another good solution for dealing with version spaces is offered by Utgoff (1986) which proposes an inductive bias to predict the more general and more specific hypothesis using the set of presented examples. This approach is probabilistic (which may not be our case as we work with a limited number of examples: 5-15) and tries to solve unseen examples; still, we appreciate the idea of hypothesis space prediction in the identification of a disjunction, and by thus of a contradictory situation.

The *split and merge* technique (Hong and Tzeng, 1999) has good results in learning disjunctive concepts. This strategy indentifies contradictions and split the boundary sets: more specific and more general hypothesis and by thus it split the version space. From this technique, we take the contradiction identification. We don't go further with the split, because in our human supporting strategy, we only intend to identify such situations and give feedback to the user (we give suggestions, not solutions).

b) Decision Trees

A decision tree is a tree represented classifier. In a decision tree, a node can be a leaf, in which case it is a class assignment, or an intermediate node, in which case, the node is the representation of a test (identifying the presence of a value for a given attribute). The whole path from the root to a leaf is a conjunction of tests which describes a class (or a concept). Usually, a class is given by the union of a set of leaves, and by thus the description of a concept is disjunctive. A decision trees receives a set of training instances (positive and negative) and build a decision tree using a greedy strategy (Quinlan, 1993), selecting the attribute that gives the best grain ratio at the current moment. If we want to use decision trees for learning conjunctive concepts, we should probably consider the rules derived from the tree and check if they contain only conjunctions of tests. Compared with version spaces, this classification method is non-incremental, i.e. the tree receives from the beginning the whole set of examples and build the classifier according to the highest gain ratio. The incremental versions of decision trees receive instances and rebuild the tree in a greedy manner; for instance, In STAGGER, the incremental induction approach proposed in (Schlimmer and Granger, 1986), the initial set of features have associated some weights that will change in order to reflect new instances introduced in the classifier. For this, it is computed the expectation of the membership of an instance to a class taking into account the previous instances (ibid.).

The decision trees are useful in learning concepts when given their human readability. The readability is given by the linear and monotonic representation of the decision trees. But, let's see first how hierarchies of concepts are represented in the human mind and then how to use this representation in the tree-based classifiers.

Tree representation in the human mind

Humans learn concepts by analogies (Johnson and Pearson, 1984). This process of forming a word which denotes a concept includes 5 steps: seeing, discussing, using, defining and writing. This form of learning a word can be used in our project in learning words by analogies (select new attributes to define the same domain). A word can identify more then one object, as claims Keil (1979) who defines the over-extension (this procedure allows using a word for a wider range of objects) and the under-extension. Both types of extensions are involved in the human usage of concepts (ibid.). This claim gives credit to the use of generalization and specialization in learning a concept (searching in a hypothesis space). Keil defines also the meta-cognition as the ability to look for the own cognitive skills. Very young people don't have a very developed meta-cognition compared with adults, as they have the tendency to classify according to concrete features (ibid.). The ability to classify using abstract features is developed after the age of 13. This finding can be useful for our project, as a possible target group is ranged from 11 to 14 years and it's useful to know that we need to combine both abstract and concrete features in learning the description of a concept.

In the hierarchical representation of concepts proposed by Keil, the concepts can be differentiated by predicates applied to their properties; these properties act like constraints and transform the concept representation in truth-trees which are monotonic and linear. Keil starts with the hierarchical representations and continue with predicate trees and ontological tree, giving constraints in representing concepts. Keil introduces also an interesting claim, according to which complex concepts or natural concepts have the visual representation of the letter M or W.

Similarly with truth-tree, predicate trees are used in representing queries with at least one path from root to the leaves (Paolino et all., 2007). This representation might be extended for learning concepts (the identification of a class/concept in a decision tree is built in the same way: identify at least one complete path). If there are more then one path, then the same leaf can give the assurance of a conjunctive concept; otherwise, we have a disjunctive representation of the identified concept (ibid.).

As we see, the human learns linear and monotonic trees for representing concepts. A similar representation of a decision tree can be of a real use in supporting the learning from examples using this tree representation.

We think that ML can be a support for the human learner; the learner is a skill developer using ML techniques to build a world around a given concept.

Before seeing related work, we present another technique that has potential for supporting the learning from examples.

c) GLD (Generalized Logic Diagrams) in the AQ algorithm

AQ21 (Wojtusiak, Michalski and Pietrzykowski, 2003) is a pattern discovery algorithm with a visual representation: GLD (Generalized Logic Diagrams). First, we present how GLD are build:

- select the set of attributes and their values;
- draw a table with one row and one column;
- select a set of attributes for the rows and the rest of attributes for the columns;
- take the number of values for the first attribute for rows and split the row into a number of rows equal to the specified number;

- for any row repeat the process with the second attribute and so on till all the set of attributes is completed;
- for columns repeat the same procedure used to split the rows.

By consequence, any cell is a collection of attributes and values. That's why the GLD can represent disjunctive concepts.

The utility of AQ21 is also reflected in:

- learning patterns (rules) with exceptions; so, it can learn disjunctions as well;
- working with irrelevant attributes;
- working with unknown values for attributes.

The last two abilities of AQ gives credit to working with incomplete descriptive (attribute-value) pairs for the training set of examples. For this reason, AQ21 (inductive pattern discovery algorithm) can be of great use when the description is not very clear (like in the case of complex represented concepts). Being so complex, AQ21 (GLD) can be used instead of decision trees when working with disjunctive concepts. Like decision trees and version spaces AQ21 has the advantage of offering a visual representation of the rules discovered. Unlike the version spaces, AQ21 is a non-incremental algorithm for pattern discovery, the same as the basic C4.5 for decision trees (or CN2, ID3 – Quinlan, 1993). However, AQ21 can deal with more complex situations then both previously presented algorithm (candidate elimination algorithms and inductive decision trees): AQ21 can represent irrelevant/missing attributes and irrelevant/missing values. In addition to this, AQ21 can identify more correct rules then the induction decision trees (Wojtusiak et all., 2006).

But, let's come back to our previous questions regarding ML techniques: *how to represent a disjunctive concept and how to recognize a conjunctive concept*? We need to investigate the potential of dealing with both types of concepts, as we are interested to recognize disjunctions and to accept conjunctions in the final description of a complex concept. In a similar way as the decision trees, GLD are likely to learn disjunctive rules and rules with exception (any exception introducing a disjunction by the means of a negation). Still, if we keep our comparison with decision trees, GLD are more likely to notify the presence of a conjunctive rule: if we redraw the GLD to group attributes we can observe that a conjunction fills a very clear shape: a *rectangle*. Any exception in such a rectangle will give credit to a disjunction. So, from this perspective, AQ21 could work better in recognizing conjunctive representations. However, there is no clear way for redrawing the GLD, but this makes this method as difficult as the redrawing of a decision tree. But, in the case of a decision tree, the rules representation in FOL – first order logic (DNF – disjunctive normal form, or CNF – conjunctive normal form) can give an answer to the conjunctive representation of a rule.

A visual representation of a rectangle in a GLD doesn't guarantee the presence of a conjunction. It can as well announce the presence of a disjunction. We can suppose, after drawing a small number of cases (attributes-value pairs represented in a GLD), that the rectangle should have one even border (even number of rows in one border); but this should be tested farther.

Our recent experiments showed that the redrawing of a GLD for a disjunctive pattern can produce a filled rectangle. We observed that a disjunction can be recognized when a larger column (containing sub-columns) is only partially filled.

For instance, when we learn disjunctive concepts represented as:

(age>=30, land=yes, education=no) or (age>=30, land=no, education=yes)

For the above example we have the following set of attributes and values: (age={>=30, <30}, land={yes, no}, education={no, middle, high}, place={village, town}). The disjunction hidden in a rectangle looks like in the table below (table 3):

		land=yes			Land=no		
		Edu=no	Edu=middle	Edu=high	Edu=no	Edu=middle	Edu=high
age>=30	place=village			+	+	+	+
	place=town			+	+	+	+
age<30	place=village						
	place=town						

Table 3. Example of GLD (the presence of a rectangle doesn't guarantee the conjunctive representation)

Another advantage of AQ21 is given by Vojtusiak et all. (2006): a GLD can represent compound and hierarchical attributes (hierarchies of values for attributes). In this case, there is still a question to be answered: how can we use this type of structure to deal with disjunctions? A possible solution can be to use the same strategy as for version spaces: replace a contradictory value (that introduces a disjunction) with a more general one (see the proposal section). But, will it work all the time? We intend to investigate further the potential of AQ21 with hierarchical attribute in the representation of disjunctive concepts.

AQ21 (or earlier inductive algorithms from the same family) is not connected to education so far. So, we think we aim to investigate the potential of this algorithm for using it in educational purposes, which is the case of our project.

ML algorithms use the simple representation of concepts. We are interested to know how to support the learning of concepts with more complex representation. That is why we analyse also the effect of typicality on algorithms. We can add therefore a question to our work: *how to use ML for supporting the learning of typicality?* Trying to answer to this question, we need to see how algorithms can deal with the graded structure of concepts.

4. ML and typicality of concepts

In this section we investigate the potential of ML algorithms to represent the typicality of concepts. As we have already seen in Rosch (1999), prototypes are the best view on representing natural concept (typicality is important in concept learning). Human concept are complex, and definitely not “crisp”, as they are in ML. The "flexible" concepts don't have clear definitions and are also context-dependent (Vojtusiak et all., 2006). Michalski proposes a two-tiered (TT) concept description for "flexible" concepts. The TT representation has two components:

1. *BCR* or *Basic Concept Representation* which covers almost any instance of a concept. The basic representation contains the basic properties of the concept, the most typical, common meaning. By this view, we are close to the classical set of attributes representation.

2. *ICI* or *Inferential Concept Interpretation* which covers exceptions and its representation is a set of inferential rules.

ICI treats exceptions, border-line cases and context dependency (for this purpose ICI uses its inferential rules to identify border cases). By this property of solving unusual cases, we can say that ICI tries to represent the natural, flexible concepts, where the apparent exceptions are also instances of a concept.

Flexibility of definition and context dependency is suitable to understand and represent in the computer the typicality of concepts. However, neither of the two components of the TT concept description provides a definition for a prototype; prototypes includes only the most representative instance of a concept and instances are grouped according to their degree of similarity to the most representative instance (Rosch, 1999).

The two-tiered representation can justify a concept learning procedure, introduced and developed earlier by (Stepp and Michalski, 1986), called *conceptual clustering*:

Conceptual clustering is an inferential concept learning strategy that searches for conjunctive concepts determined by a reference criterion. The procedure finds first the classifying attributes and applies inference chains to associate a class to an instance. The inferential search for a description seems to us justifiable by the two-tiered representation: a classifying attributes set (the BCR, as mentioned before) and a rule-set for classifying exceptions (the ICI).

The TT representation was used in relation with the AQ15 (Michalski, 1992) who proved that the search using the TT description can provide better classifications. TT was also used in relation with the decision trees representation as presented in (Bergadano et all., 1992). We think that TT has a potential of being used also with version spaces (Mitchell, 1978); we suppose that the generalization/specialization procedures needed in a version space can be provided by the context in which an instance appears.

In this section we investigated the AI representation of concept and a learning algorithm: the two-tiered concept representation and the conceptual clustering. The two-tiered representation of concepts supports us to understand and represent typicality in ML.

Before continuing, we think that it's worth mentioning briefly the flexibility of defining concepts as it is presented in (Winston, 1975). We do this, because the *natural concept* that we've investigated so far allows the definition flexibility in choosing the characteristic attributes (definition accepted by the graded-structure theory). Winston mentions the importance of learning conceptual descriptions with modal attributes; he uses the must/must-not, can/cannot properties of a characteristic attribute. Using the mentioned defining values, Winston (1975) allows concepts to choose the presence of an attribute or its absence. A similar conceptual significance can be found in how 3-valued logics work. Turner (1985) mentions the ***3 – valued logics*** which accept a third truth value, *u* (="*undecided*", partially ignored). As the purpose of our review is not related to logics (except for deontic logic, in chapter 4), we won't insist on them, but we synthesize our findings and suppositions of how different types of logic accept the *must/can* values in the table below. In the following table we try to relate different types of logic with natural concepts; we want to underline the possible connection between logics and concept description:

Value	**Dynamic Deontic-Epistemic logic**[1]	**Modal logic**$_1$	**3-valued logic**	**Concept description**
Must/ Must-not	Represents the *obligation/forbidden actions* of an agent (students are in this case agents involved in a learning environment)	The agents act according to the *necessity* modal operator	Classical truth values: true/false (signal for the case when a values is given to/misses from an attribute)	The learner receives *positive/negative instances* which characterize the concept
Can/ Cannot	Reflect the *permission*s into an agent's actions	Attributes can appear or not in the description of the concept (flexibility in defining the concept)	A value for *ambiguous* cases (unknown or undefined attributes/values)	The classification of *unclear cases* (borderline instances of "natural concepts")

Table 4. Logics in Concept Learning (possible similarities)

5. AI in education

ML techniques were already applied in education to simulate the learning process or to offer suggestions. We will see the previous approaches of ML that can be relevant for our project. We will then two approaches, centred on human learning this time (focus on the Socratic tutors), in the chapter 4. But, first let's see how ML techniques were used in education.

Previous approaches of ML in AI

First we will see how are used the ML techniques that we consider suitable for supporting the human in the learning process:

1. *Version spaces* were successfully used in learning physics in the form of story building. The idea is to represent a states space and enrich the hypothesis space by adding a new instance. The authors don't mention explicitly this method, but they mention the space extension and the generation of states of the space. They introduce also the "explicit manipulation of state space" (Pearce and Luckin, 2007). Unless the candidate elimination algorithm, Pearce and Luckin focus on the discovery of an instance (the whole scenario generating states that characterize a complex instance; e.g. in order to learn the concept of gravitation, a student can learn a story in which a glass of water is turned down, the water is on the floor and not in the glass anymore). This technique can be successfully used in collaborative learning, when students build the story by adding new sentences to the scenario.

2. Decision trees are successfully used in diagnosis system (Ordonez, 2006). For instance, decision trees are used to diagnose healthy and diseased arteries. Decision trees are also used in nursing education (Kokol, 1999); this strategy supports decision making.

---

[1] *Modal logic* works with modalities; it defines modal operators, which act as possibility and necessity (Lutz C., Wolter F. and Sattler U., 2008).
Deontic Logic analyses the obligations and permissions of agents in a system (van der Torre, L. and Hansen J., 2008).
Dynamic Epistemic Logic defines a knowledge system, with agents and actions, and changes of states. A state is a possible world (van Ditmarsch, H. and van Eijck J., 2008; Fagin R., et all.., 2003).

3. Logic representations of human learning: FOIL and KBG (Emde, 1995) which use conceptual clustering; the system is user-oriented as the user has the possibility to analyse the rules and to adapt the parameters of the system. The rules are explained by the system, in order to be understood by the human user. The system receives as input case-oriented representations in the form of objects which are conjuncts of ground literals and rules which represent the domain theory. FOIL uses a first order logic (FOL) representation to build a single clause using a general-to-specific strategy. Any positive example adds information to the information gained by the system: $I(T_{i+1})=-\log((T_{+(i+1)} + c)*(|T_{i+1}| + c))$. The need and importance to study a disjunctive representation of concepts is motivated by the fact that complex-defined concepts may be difficult to be characterized in terms of (attribute, value) pairs. In this case, a disjunctive representation is possible because of a several causes:

- a poor descriptive language: not enough attributes or values in the vocabulary;
- a wrong classification of an example (wrong class assignment);
- a wrong attribute/value assignment to an example (ibid.).

We need to mention again that *contrary to the tendency in ML field*, *we are not interested to know how to learn concepts using ML techniques, but how to use ML to support the learning process*. **ML should provide a guidance, not a solution**, in a similar way the Socratic tutors work. So, let's explain what a Socratic tutor is, why we need this kind of learning technique in our approach, and how other people adapted it in the context of using artificial intelligence in education.

Socratic tutors implement a pedagogical strategy called, ***Socratic method***. The teacher that uses the Socratic method guides the students by asking questions; the teacher takes into account the background and the answers of the student and helps the student reflect on the own answers and find the target concept alone (Trella et.all, 2006). This method was used in association with images, by Trella (2006), for learning botany concepts, more exactly European forestry species. Trella presents to the students images representing trees and ask them to recognize the species. The knowledge related to the images is maintained into a conceptual domain in the form of a hierarchy of concepts. The learning strategy involves that the teacher asks and receives questions: if the student fails, the teacher reformulates the question giving a hint; there is not a big fail if the student recognize a concept with the same parent in the hierarchy (hypernym) as the target concept (ibid). This approach is very useful for us, as it is using, the same as we want, a guided dialogue strategy based on the use of images for learning concepts.

Before continuing, we present another approach, interesting for us, which is based on the idea of thinking of concepts inside a dynamic learning system (this is important for the "natural concepts" which develop during the learning process, not having the defining attributes from the beginning). Sweeney and Sterman (2007) introduce the "system thinking" and define the notion of ***system concepts***. They are interested to understand complex dynamic systems (e.g. climate change) and build a system based on two components: 1)conceptual knowledge (the use of structures and recurrent patterns of behaviour, that the student is required to recognize) and 2)reasoning skills (the changing of behaviours and the recognition of *homologies* – recognize the same problem in another context; this fact is related to the transfer of skills[1]). Contrary to the previous approach, Sweeney and Sterman don't focus on learning terms (concepts), but on how to describe the *feedbacks*. As the feedback is important in the learning process, they investigated different results in how young and adults recognize feedback: youths

[1] Skills transfer involves the translation of task-specific skills into discrete skills, the mapping of logical source objects to the logical target-tasks objects by extracting rules defining source skills (Torrey, Shavlik, et al., 2006).

pay attention to feedback because they are more likely to recognize close entities than adults who recognize distant entities based on linear relation of *causality* and recognize the positive feedback.. Another finding of the experiments made by Sweeney and Sterman (2007) involves the way the different categories of users define learned concepts: youth (middle-school pupils) explain the concept using the target term (the concept itself) and pronouns, while adults (science and non-science teachers) use definitive features of the target concept; the experimenters justify students' explanations by youths' ability to recognize reinforcing feedback (e.g. population growth related to the number of births) and the teachers' tendency to find characteristics of a concept because of the ability of adults to recognize positive feedback. This approach concluded with question related to the use of feedback in the reasoning process (ibid). If we reformulate their question for our purpose, we can ask: *Why not incorporate feedback into AI support*? We believe that the feedback received in a visual representation helps the learner to articulate faster the concept. Therefore, let's explain how the visualization and reasoning work together:

Interpretation and reasoning in visual representations

The learners use images which contextualize concepts, and that's why we need to analyze the readership skills required by the classification process as part of the learning process. Learners cluster objects according to similarities and usually they do this manually, by a drag and drop procedure. This process of *natural clustering* is described in the VITE visual interface, presented in (Hsieh and Shipman, 1991) , interface which allows users to manipulate structured data. In the VITE interface, Hsieh and Shipman observe the graphical mapping, which they call perspective, between specific attributes of data and their graphical properties; the mapping between objects and representations is important for reuse and redesign of data. VITE allows users to edit new mappings, using of semantic attributes, limited enough to insure the differences (ibid.).

If we are focused on clustering images as a learning process, we must pay attention to the creative part of the natural clustering. That's why, we have to think of the classification strategy as a *knowledge building process*. (Scardamalia and Bereiter, 1994) introduces the CSILE (computer-supported international learning environments) to underline the importance of building a learning *discourse*; in CSILE, in a similar way to the Socratic dialogue, the teacher guides the learners, but students can influence their evolution by adapting the environment. From this point of view, Scardamalia and Bereiter define the first-order and the second-order environment of the *knowledge-building community*: the first-order environment is focused on asymptotic learning, where the learning is student-centred and the learning doesn't influence the environment (an individual perspective of learning), while in the second-order environment is adapted to changes made by different users (a collaborative, constructive learning). CSILE uses both environments and store the dialogue between an individual student and the teacher and in the same time allows students to share their knowledge.

Apart from the individual and group aspect of the learning process, the visual representation of concepts is also an evaluator for the learning process. According to Petre (1999), which analyzes the visual representation carried out by a novice and an expert in programming, more experienced people are able to identify the missing element of a description in the graphics that accompany the text (she identifies the ability of experts to use the secondary notations. By secondary notation it is understood the textual comments which explain the code of the program). We hope that students will be able to evaluate themselves in a creative way, by classifying annotated images and also to articulate the concept they learn.

Another form of visualizing the concepts that a student learn is a ***concept map***.

Concept maps

Concept maps, as mentioned in (Kremer, 1996), are graphs of nodes labelled with text representing the "concept" and arcs, labelled with a relationship type (arcs represent relationships between nodes). Both nodes and arcs can have visual attributes (i.e. shape, colour etc.). Concept map can be used individually or for group applications. The individual uses a concept map to "convey non-linear information structures" (Kremer, 1996). Concept maps (Kremer, 1996) are graphs of nodes labelled with text representing the "concept" and arcs, labelled with a relationship type (arcs represent relationships between nodes). Both nodes and arcs can have visual attributes (i.e. shape, colour etc.). A similar definition is given by Sowa (2000) for conceptual graphs: "A *conceptual graph* is a bipartite graph that has two kinds of nodes, called *concepts* and *relations*". Every arc in the graph links a concept to a conceptual relation (Sowa, 2000:477).

Concept map can be used individually or for group applications. The individual uses a concept map to "convey non-linear information structures" (Kremer, 1996). The paper focuses on the multi-user use of concept map in relation with web documents, as concept maps can describe a collection of documents; the Web creates a partnership model, and by consequence it can be used by multiple users.

Concept maps are important for multi-user interactive documents for several reasons:

1) each user may use/owe his own concept map;

2) all users update the shared concept map on the server;

3) server will broadcast the updated concept map to all participants;

4) each user can see the shared concept map and concept maps of all participants;

5) each user can adapt the own concept map.

Concept maps are important in general, for individual and multi-uses because they:

1) can convey structures: sequences, hierarchies, cycles, containments etc.;

2) make explicit the relationships between nodes (visually distinguished);

3) nodes can be visually distinguished (ibid.).

Kremer (1996) presents two structures similar to concept maps: constraint graphs and conceptual graphs. *Constraint graphs* are used to visually describe a formalism by the means of levels that may contains only three basic objects: 1) nodes, 2) arcs, 3) contexts; the basic types can determine a lattice of subtypes, in which any s*ubtype* introduces *new attributes* and predicates (that extend the attributes) to constrain the graph to a particular formalism (predicates can be used for disambiguation). The other type of structure is the conceptual graph, which uses two components: a graph (modelling nodes and relationships between them) and a semantic net (specifying the type lattice: the hierarchy of types). By consequence, constraint graphs are extensions of conceptual graphs.

Concept maps are used as formal and informal representations. Concept maps can be applied in brainstorming sessions, where users can compare, arrange, contrast and group ideas (ibid.).

Kremer claims that concept maps are a "intuitive form of knowledge representation". If we agree with this hypothesis we may also consider the tacit dimension (Polanyi, 1967) in the learning while using concept maps. This fact can be sustained by the affirmation of (Lin and Sun, 2008) who argue that concept maps are used for internalization and externalization of knowledge. Therefore, we can ask: *Why not to use the knowledge representation to contribute to the tacit knowledge (such as by using tag clouds and concept maps) and to reflect changes in the learning process (or, in other words, modify own concept map to integrate within a learning scenario proposed by a teacher)?*

Conceptual maps are useful for their property to reflect different understanding of concepts. Concepts are perceived in different ways by different students and these differences create the concept boundaries (Lin and Sun, 2008). *Concept boundaries* are used for further learning: students compare and contrast their understanding and adapt it so they reach a common understanding. Students are also given the possibility to re-evaluate their own concept maps and build *integrated concept maps* in order to visually perceive the conceptual differences; after they see the integrated concept maps, students will be asked to add or delete nodes in order to reflect their new understanding of concepts and relations between concepts in the map. It's important that the teacher requires students to make only necessary changes in the concept hierarchies so that there will be a mapping between the understanding of different students. Results presented in (Lin and Sun, 2008) showed the improvement of self-awareness (the ability to reason about the own work), but only 25% of students were aware that this method of learning helped them to improve the recognition of their own faults. Lin and Sun conclude that *meta-cognition and concept-mapping work together by breaking concept boundaries and improve creativity*.

We hope that ML will be a great support for the human learner if it will allow the reasoning process to take place in a visual environment.

The previous studies shows that the techniques that we are interested in have a certain potential for supporting the learning. We will need to show within this PhD work how to adapt and combine AI for supporting the human learn concepts, and acquire skills as well, and more exactly, on how to learn concepts from images.

3. Conclusions and future work

So far we investigated the potential of some ML techniques for supporting the learning of concepts from images. We investigated concepts and the tacit dimension of knowledge and learning. We will focus now on building case studies, building and testing a learning software that will implement the architecture of our proposal (which will be presented in a future report). We hope that our work will answer the question of using AI in supporting not only students learn concepts from images, but also adults in more difficult conceptual challenges like business management or knowledge acquisition (as knowledge implies the use and understanding of concepts).

4. Acknowledgements

I have to thank to my supervisors Dr. Zdenek Zdrahal and Dr. Paul Mulholland for their continuous support and the learning opportunities they offered.

5. References

Anderson, J. R. (2000). *Learning and memory, an integrated approach*, second edition, John Wiley and Sons, Inc., New York

Bergadano, F., et all. (1992). *Learning Two-Tiered Descriptions of Flexible Concepts: The POSEIDON System* in *Machine Learning 8(1): 5-43*, pp. 5-43, Klever Academic Publishers, Boston, Manufactured in The Neederlands

Berry, D.C., Broadbent, D.E. (1984). *On the relationship between task performance and associated verbalizable knowledge*. The Quaterly Journal of Experimental psychology, 36A, 209-231

Borger, R., Seaborne, A.E.M. (1982). *The Psychology of Learning*, Second Edition, Penguin Education, Richard Clay

Bourne, L. E.Jr. (1970). *Knowing and using concepts* in *Psychological Review*, Vol. 77, No. 6, pp. 546-556

Cook and Brown (1999). *Bridging Epistemologies: The Generative Dance Betweeen Organizational Knowledge and Organizational Knowing*, in Organization Science, vol. 10, no. 4, july-august 1999

Emde W. (1995). *Machine Learning: Case Studies of an Interdisciplinary Approach*, in *Learning in Humans and Machines* (Eds. Reimann, P. and Spada, H.), Pergamon, Elsevier Science Ltd., Oxford, UK

Fagin R., et all. (2003) – *Reasoning about knowledge*, MIT, London

Fu, L.-M. and Buchanan, B. G. (1985). *Learning Intermediate Concepts in Constructing a Hierarchical Knowledge Base*. In *Proc. of the 9th Int. Joint Conf. on AI*, pp. 659--666.

Hassan-Montero, Y. and Herrero-Solana, V. (2006). *Improving Tag-Clouds as Visual Information Retrieval Interfaces*, International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006, Merida, Spain, October 25-26, 2006

Higgins, S., Packard N. (2004). *Meeting the Standards in Primary ICT. A Guide to the ITT NC*, Routledge Farmer, New York

Hong, T. and Tzeng, S. (1999). *Splitting and merging: an approach to disjunctive concept acquisition*, in *IEEE Trans. Knowl. Data Eng. 11(5): 813-815 (1999)*

Hsieh, H.-S. and Shipman, F.-M. III (2000). *VITE: a visual interface supporting the direct manipulation of structured data using two-way mappings* in *Proceedings of the 5th international conference on Intelligent user interfaces*, New Orleans, Louisiana, United States, ACM.

Johnson, D.D. and Pearson, P.D. (1984). *Teaching reading vocabulary*, 2nd edition, Hold, Rinehart and Winston, USA

Kang, S., Scharmann, L. C. and Noh, T. (2004). *Reexamining the Role of Cognitive Conflict in Science Concept Learning*, in Research *in Science Education 34: 71-96*

Keil, F. C. (1979). *Semantic and Conceptual Development. An Ontological Perspective*, Harvard University Press, Cambridge, Massachussetts

Klausmeier, H., Allen, P. (1978). *Cognitive Development of Children and Youth. A longitudinal study*, Academic Press, New York

Kokol, P. et all. (1999). *Decision Trees - A CIM Tool in Nursing Education*. In Proceedings of HPCN Europe 1999, pp. 951-958

Kremer, R. (1996). *Toward a Multi-User Programmable Web Concept Mapping "Shell" to Handle Multiple Formalisms*, in *Proceedings of 10th BANFF Knowledge Acquisition for Knowledge-Based Systems Workshop,* Vol. 2, Banff, Alberta, Canada, Nov. 9-14, 1996, Gaines, B.R. and Musen, M. (Eds.)

Laurillard, D. (2002). *Rethinking University Teaching. A conversational framework for the effective use of learning technologies*, 2nd edition, Routledge Falmer, London

Lieberman, H. (1987). *An example-based environment for beginning programmers* in Lawler, R. W. and Yazdani, M., 1987 – *AI and Education, Vol. 1: Learning Environments and Tutoring Systems*, Ablex Publishing, New Jersey

Lin, G. Y-M.,and Sun, .S.J. (2008). *Breaking concept boundaries to enhance creative potential: Using integrated concept maps for conceptual self-awareness*, in *Computers & Education, no. 51, pages 1718-1728*, Elselvier Ltd., 2008

Lutz C., Wolter F. and Sattler U. (2008) – *Modularity in Logical Theories and Ontologies*, courses from the 20th ESSLLI Summer School course, Hamburg, Germany, 4th-15th August 2008

Mitchell, T. M. (1977). *Version spaces: A candidate elimination approach to rule learning*, IJCAI 5, pp.305-310

Murphy, G. (2002). *The Big Book of Concepts*, The MIT Press, Cambridge, Massachussetts

Nonaka, I. (1991). *The knowledge creating company*, Harvard Business Review, 1991

Nonaka, I. (1994). *A Dynamic Theory of Organizational Knowledge Creation, Organization Science*, Vol. 5, No. 1, (Feb., 1994), pp. 14-37

Ordonez, C. (2006). *Comparing Association Rules and Decision Trees for Disease Prediction*, in *Proceedings of the international workshop on Healthcare information and knowledge management*, Arlington, Virginia, USA

Paolino, L. et all (2007). *The Predicate Tree – A Metaphor for Visually Describing Complex Boolean Queries*, in *Advances in Visual Information Systems 9th International Conference, VISUAL 2007* Shanghai, China, June 28-29, 2007

Pearce, D. and Luckin, R. (2007). *The Principle of State Expansion in Task State-Space Navigation*, in *AI in Education*, ed. Luckin, R. et all, IOS Press, 2007

Petre, M. (1995). Why *looking isn't always seeing: readership skills and graphical programming* in *Commun. ACM 38(6): 33-44*.

Polanyi, M. (1967). *The tacit dimension*, Routledge & Kegan Paul London U.K.

Pritchard, A. (2005). *Ways of Learning. Learning Theories and Learning Styles in the Classroom*, David Fulton Publishers, Abington, Oxon, Great Britain

Quinlan, R. (1983). *C4.5: Programs for ML* by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993

Reber, A. S. (1993). *Implicit Learning and Tacit Knowledge. An Essay on the Cognitive Unconscious*, in *Oxford Psychology Series* No. 19, Oxford University Press, Calendron Press, New York

Reyhan, A. and Pinar, Y. (2007). *Learning Costumer Preferences using Semantic Similarities*, in *Proceedings of AAMAS07*

Rieder, A. (2003). *Implicit and explicit learning in incidental vocabulary acquisition*, in *Views*, University of Wien

Rosch E. (1999). *Reclaiming concepts*, in in R. Numez & W. J. Freeman Editions, *Reclaiming cognition: the primacy of action, intention and emotion*

Rosh, E. (1978). *Principles of Categorization*, in Editors Rosh, E., Lloyd, B.B., 1978 – *Cognition and Categorization*, John Willey & Sons, New York

Scardamalia, M. and Bereiter, Carl (1994). *Computer Support for Knowledge Building Communities,* http://carbon.cudenver.edu/~bwilson/building.html

Schlimmer, J. C. and Granger, R. H. (1986). *Incremental Learning from Noisy Data*, in *ML* 1: 317-354, 1986, Kluwer Academic Publishers, Boston - Manufactured in The Netherlands

Shapiro, A. and Niblett, T.B. (1982). *Advances in Computer Chess 3* (ed. Clarke, M.R.B.), pp. 73–92 Pergamon, Oxford

Smith E., Medin D. (1981). *Categories and Concepts*, Harvard Uni Press, Cambridge, Massachussetts

Sowa, J. F. (2000). *Knowledge Representation. Logical, Philosophical and Computational Foundations*, Brooks/Cole, Thomson Learning, Pacific Grove, USA

Stepp, E.R. and Michalski, R.S. (1986). *Conceptual clustering: Inventing Goal-Oriented Classi_cations of Structured Objects*, in *ML. An AI Approach*, Volume II, (eds.Michalski, R.S., , Carbonell, J.G. and Mitchell, T.M.), Morgan Kaufman Publishers, Inc., Los ALtos, California

Sternberg, R.J. and Horvarth, J. A. (1999). *Tacit knowledge in professional practice*: researcher and practitioner perspectives, Mahwah, N.J; London: Lawrence Erlbaum, 1999

Sussmann, G.-J. (1975). A *Computer Model of Skill Acquisition*, Americal Elsevier Publishing Company, Inc., New York

Sweeney, L. B. and Sterman, J. D. (2007). *Thinking about systems: student and teacher conceptions of natural and social systems*, in *Systems Dynamics Review, nr. 23, vol. 2/3, pages 285-312*

Torrey, L., J. Shavlik, et al. (2006). *Skill acquisition via transfer learning and advice taking*, Heidelberg, D-69121, Germany, Springer Verlag

Trella, M., Conejo, R., Guzman, E. (2006). *A Web Based Socratic Tutor for Trees Recognition*, in *Adaptive Hypermedia and Adaptive Web-Based Systems. International Conference, AH 2000. Proceedings*

Turner, R. (1985). *Logics for Artificial Intelligence*, Ellis Horwood Limited, England

Utgoff, P. (1986). Shift *of bias for inductive concept learning,* in *Machine Learning: An Artificial Intelligence Approach, Volume II*. (Ed. R.S. Michalski, J.G. Carbonell, and T.M. Mitchell), Morgan Kaufmann, California

van der Torre, L. and Hansen J. (2008) – *Deontic Logic in Computer Science*, courses from the 20th ESSLLI Summer School course, Hamburg, Germany, 4th-15th August 2008

van Ditmarsch, H. and van Eijck J. (2008) – *Dynamic Epistemic Logic*, courses from the 20th ESSLLI Summer School course, Hamburg, Germany, 4th-15th August 2008

Weller, M. (2007). *Virtual Learning Environments. Using, choosing and developing your VLE*, Routledge, Oxon, UK

Winston, P. H. (1975*). Learning structural descriptions from examples* in *The psychology of computer vision*, Winston, P. H. (Ed.), Mc-Graw-Hill, New York, 1975

Witten, I. H. (2005). *Data mining: practical ML tools and techniques*, Elsevier Ed., Amsterdam, 2nd edition, 2005

Wojtusiak, J., Michalski, R. S. and Pietrzykowski, J. (2003). *Multiple Pattern Discovery via AQ21. A Brief Description of the Method and Its Novel Features*, in *MLI 06-2, P 06-2*, George Masson University, June 5, 2006

Wojtusiak, J., Michalski, R. S. and Pietrzykowski, J. (2006). *The AQ21 Natural Induction Program for Pattern Discovery: Initial Version and its Novel Features* in *Proceedings of the 18th IEEE International Conference on Tools with AI*, pages 523-526, 2006, ISBN ~ ISSN:1082-3409 , 0-7695-2728

Zhu, X., Zhu, D., Lee, Y., Simon, H. A. (2003). *Cognitive Theory to Guide Curriculum Design for Learning from Examples and by Doing*, in *Journal of Computers in Mathematics and Science Teaching*, 2003, 22 (4), pages 285-322

**** http:// www.bidgemaneducation.com, (consulted in 2008). *Bridgeman Art Library, London*

**** http:// http://www.cs.waikato.ac.nz/ml/weka/ *(consulted in 2008). WEKA tool which implements ML algorithms for data mining*

**** http://www.standards.dfes.gov.uk/schemes2/citizenship/?view=get, (consulted in 2008). *The history learning curricula: Citizenship and history, Unit 12 - Why women and some men have to struggle for the vote in Britain? What is the point of voting today*?

*** http://www.wordenet.princeton.edu, WordNet (version 3.0, consulted in 2008)

List of tables

List of figures