

ABEL, Mara; MASTELLA, Laura Silveira; SILVA, Luis Alvaro Lima; CAMPBELL, John; De ROS, Luiz Fernando. How to Model Visual Knowledge: a study of oil-reservoir evaluation expertise. In: 15TH INTERNATIONAL CONFERENCE AND WORKSHOP ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, 2004, Zaragoza, Spain. Conference Proceedings with LNCS. 2004.

How to model visual knowledge: a study of expertise in oil-reservoir evaluation

Mara Abel¹, Laura S. Mastella¹, Luís A. Lima Silva¹, John A. Campbell², Luiz Fernando De Ros³

¹ Federal University of Rio Grande do Sul - UFRGS
Instituto de Informática - Porto Alegre, Brazil
{marabel, mastella, llima}@inf.ufrgs.br

² University College London
Dept. of Computer Science - London, UK
j.campbell@cs.ucl.ac.uk

³ Federal University of Rio Grande do Sul - UFRGS
Instituto de Geociências - Porto Alegre, Brazil
lfderos@inf.ufrgs.br

Abstract. This work presents a study of the nature of expertise in geology, which demands visual recognition methods to describe and interpret petroleum reservoir rocks. In an experiment using rock images we noted and analyzed how geologists with distinct levels of expertise described them. The study demonstrated that experts develop a wide variety of representations and hierarchies, which differ from those found in the domain literature. They also retain a large number of symbolic abstractions for images. These abstractions (which we call visual chunks) play an important role in guiding the inference process and integrating collections of tacit knowledge of the geological experts. We infer from our experience that the knowledge acquisition process in this domain should consider that inference and domain objects are parts of distinct ontologies. A special representation formalism, k -graphs⁺, is proposed as a tool to model the objects that support the inference and how they are related to the domain ontology.

Keywords: Knowledge acquisition, knowledge representation, expertise, visual knowledge, petroleum exploration

1 Introduction

Knowledge about the evaluation of oil-reservoir rocks is crucial in petroleum exploration, since it can substantially decrease the risks of exploration and increase the efficiency of hydrocarbon production. Most of the data relevant for geological interpretation of oil reservoirs consist of visual information that cannot be described only

through geometric components, such as size and format. Many of the aspects recognized by a geologist during the interpretation task have no formal denomination and are learnt through an implicit process during training and field experience.

The use of features without names in supporting problem-solving is a current practice in many natural domains. These objects constitute the implicit body of knowledge, also called *tacit knowledge* by Nonaka and Takeuchi [1] when referring to the unarticulated knowledge that someone applies in daily tasks but is not able to describe in words. The articulated or *explicit knowledge* refers to the consciously recognized objects and how these objects are organized. This portion of knowledge in the context of artificial intelligence is called *ontology*. Tacit and explicit knowledge should be seen as two separate aspects of knowledge and not different sorts of it. Extracting the ontology of a domain is one important objective of knowledge acquisition techniques; sometimes, it is misunderstood to be the only one. Along with the ontology, which represents the explicit part of knowledge, it is necessary to identify the tacit knowledge applied by experts and propose knowledge representations for it.

In geology, as in many other natural domains, the role of tacit knowledge in the expert problem-solving process has a dominating influence on the results achieved. Therefore, it deserves special attention in knowledge engineering. This paper describes a study in the petroleum domain that identifies what kind of tacit knowledge is applied by geologists in the evaluation of oil reservoirs, as part of the *PetroGrapher* project for the development of an expert system to support reservoir-rock interpretation. The study acknowledges and identifies differences between the cognitive mechanisms of novice and expert geologists during the rock interpretation process. Further, we discuss how these differences can influence the choice and application of knowledge-acquisition techniques.

2 Knowledge acquisition and knowledge elicitation techniques

Knowledge acquisition refers to the process of collection, elicitation, interpretation and formalization of the data regarding the functioning of expertise in a particular domain. Its objective is to reduce the communication gap between the expert or knowledge worker and the knowledge engineer, allowing the knowledge to become independent of its sources.

The main classes of knowledge acquisition techniques are described briefly below.

- Interviews, observation and protocol analysis: A grouping of many different techniques that demand direct interaction with experts. In retrospective interviews, the expert narrates a memory of how a problem was solved. This description commonly omits many crucial details. In concurrent interviewing, e.g. via observation and protocol analysis, the expert verbalises his/her reasoning *during* the problem-solving process while it is being recorded and observed. The result is more trustworthy, but the expert is usually unable to verbalize what he/she is doing when the inference requires both a high level of abstract reasoning and a low (i.e. concrete) level of sensorial activity. The collected information is commonly imperfect and needs to be complemented through further techniques.

- Classification techniques aim to identify the terms and concepts of the domain and how these concepts are organized in classes, groups or components, according to the expert. These include card-sorting and multidimensional scaling techniques [2].
- Collecting cases: A general label for all techniques that exploit recorded cases in knowledge acquisition, such as scenario analysis, event recovering, and the analysis of legacy cases for use in case-based reasoning systems.
- Extracting cause-effect relations. This includes techniques used to extract causal relations among concepts of the domain (such as evidence for conclusions, or problems and their applicable solutions). Variations on repertory grids [3], rule-extraction, knowledge graphs [4, 5] and conceptual graphs [6] belong in this class.
- Identifying the reasoning path. Problem-solving methods (PSM) [7] and inference structures [8] are graphical representations of the inference process involved in problem solving, described at an abstract (though not generic) level.

The accepted knowledge acquisition techniques are effective in revealing the ontology (explicit knowledge) underlined in expert reasoning. However, little progress on elucidating the unarticulated parts of expert knowledge has occurred, e.g. visual recognition or the integration of sensorial objects in the domain ontology. These issues are the main focus of the present work.

3 Investigation of the cognitive abilities of experts in petrography

Geological interpretation is strongly based on visual interpretation of features imprinted in rocks or landscape by physical phenomena. Since this process is improved mainly by practical experience, and not by supervised learning, geology is one of the areas in which experts develop themselves as strategic sources of knowledge.

The most familiar achievements in knowledge acquisition for geological domains are the development of the PROSPECTOR system [9]; the XEOD expert system to identify detrital depositional environments [10]; and the knowledge acquisition project SISYPHUS III, concerning igneous petrography [11]. The influence of the domain on the efficacy of the knowledge acquisition technique is significant, as was discussed by [12]. The common aspect of all these projects is that they have drawn attention to the development of dealing with visual diagnostic features and also have shown the necessity of representing the objects that support inference within a separate ontology.

In the present study, our intention has been to elucidate the cognitive process and objects that support the expertise in sedimentary petrography, by using tests similar to those found occasionally in cognitive psychology, e.g. in [13].

The investigation was conducted over a group of 19 geologists with distinct levels of expertise in sedimentary petrography. The group was selected among lecturers, undergraduate and graduate students of the Geosciences Institute of UFRGS and geologists from a petroleum company. Practical experience, instead of only theoretical knowledge, was a fundamental prerequisite for the selected group. The members of this group were first classified as novices, intermediates or experts. Novices were students or geologists who had received at least 100 hours of training in sedimentary petrography. Intermediates were geologists who used petrography as a daily tool in

their work. Experts possessed at least 10 years of experience in the subject and directly utilized sedimentary petrography for more than 10 hours per week.

The group was requested to carry out 5 different tests, based on the presentation of images from rocks, using a high-definition video system attached to an optical microscope. The first experiment was designed to evaluate long-term memory. They were requested to describe fully a thin section that had been examined one hour before. The second test requested a full description of a thin section without time restriction, in order to evaluate the richness of technical vocabulary. The third test investigated short-term memory, by requiring description of a rock image shown just previously for a very short time. In the fourth experiment, recall of a first set of pictures of common objects (animals, landscapes and people) and a second set of sedimentary rocks was requested just after these had been shown for a short time. In the fifth experiment, the geologists were asked to divide another set of pictures into subsets (i.e., classify them) and explain their criteria for this subdivision. The pictures were basically reservoir images under microscope and, in a second experiment, common scenes, such as people, landscapes, etc.

The set of experiments was conceived in order to measure the association of 3 indicators with the predefined class of expertise. The indicators were:

1. amount of significant information in the description obtained after or during image exposition: it was expected that experts knew more about the domain and that their knowledge would be expressed through the use of technical and precise vocabulary;
2. intensity of use of interpreted features, instead of features having objective geometrical properties: it was expected that experts would develop image recognition at a more abstract level than novices and would be able to demonstrate this in recognizing features that needed only a short process of inference for their identification;
3. efficiency of organization and indexing of the domain: it was expected that experts would be more effective in grouping and classifying new information related to the domain and could demonstrate this ability in experiments involving their memory.

In the results of the experiment, we found no relation between the quantities of words or even of significant words (which are the words really related to the domain) in descriptions produced by experts and by novices. On the average, experts record more efficiently the details of the rock inspected, but the overall relation is not so simple: some novices exhibit an expert level of performance and some experts perform like novices.

When considering the use of interpreted features in the description, experts clearly demonstrate a higher average, as shown in Fig 1. The vertical axis in this figure indicates how many clearly interpreted features were used in describing the rock. The general pattern was the same in the short-term memory experiments involving sedimentary rock samples: the faster the expert is requested to classify or interpret rocks, the more he/she will try to recognize diagnostic features.

Our experimental picture changes completely when another kind of rock is used, such as a metamorphic rock sample. The result can be seen in Fig 2, where no relation between level of expertise and interpreted features can be identified. The sample here includes a professor of metamorphic petrology, classified as a novice in the context of sedimentary petrography. As expected, the geologist exhibited an expert pattern of behaviour in his own domain of knowledge.

Another significant result was obtained from the experiments with the set of images. Experts show a worse memory for common photographs than the novices, but obtain excellent results when the pictures concern their domain of expertise.

Evidence for the ontological support for memorisation was demonstrated when experts were requested to organise pictures of common objects and rocks. In the classification of rock pictures, the experts utilised evidently interpreted aspects of the rock, such as the quality of porosity or the kind of cementation. The novices organised the pictures mainly in terms of colour, texture or abundance of minerals. The organisation of the domain in the expert's mind relies on aspects that make the *problem-solving process* easier or more effective, rather than the taxonomy commonly used for teaching of students.

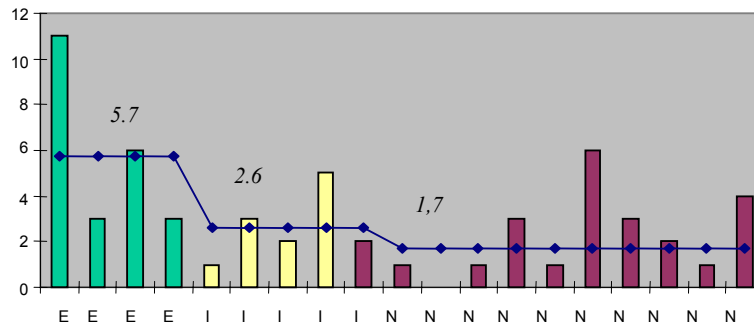


Fig 1. Relation between the level of expertise and the usage of interpreted features in information extracted from thin-section rock images by geologists in their descriptions of rock samples. The bars indicate the number of interpreted features and the line represents the median for each class (experts (E), intermediate (I) and novices (N))

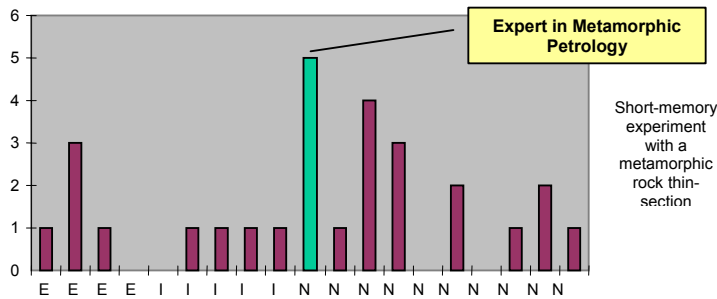


Fig 2. The same experiment as covered in Fig 1, but using a metamorphic rock sample. The relation between expertise and amount of features is not longer perceived. The highest bar refers to a professor of metamorphic petrology, who exhibits an expert behaviour in this experiment.

The experiments were able to identify that experts apply interpreted features more often than novices do in the problem-solving process. These features not only provide ontological support which improves memorisation of pictures concerning the domain

of expertise, but also make the expert's problem-solving process more effective, even though in most instances the geologist can not give them a proper name. Given the observed importance of these cognitive objects for the process of interpretation, further cognitive research to establish methods tuned specifically for acquisition and representation of such knowledge is desirable.

4 The knowledge acquisition process

The experiments showed clearly that the domain has more than one level of mental model. One, more connected with the communication process among geologists, is most easily elicited with the interviews and classification techniques mentioned in Section 2. The knowledge applied in rock interpretation, however, is mainly unconscious and can scarcely be extracted during conventional interviews.

In this project, the main source of knowledge was an expert in diagenesis of siliciclastic rocks with more than 15 years of experience in reservoir evaluation. The knowledge-acquisition process was carried out in two distinct phases. In the first phase, traditional knowledge-engineering techniques were used, such as bibliographic immersion, interviews using retrospective and concurrent protocols, and card-sorting. The main goals of this phase were the identification of the user's expectations for the system and the specification of the object hierarchy of the domain that would compound the domain knowledge in sedimentary petrography. At this phase the schemata of the knowledge base was generated.

In the second phase we tried to elicit the knowledge that actually supported the rock interpretation. The expert's reasoning in conducting the rock-sample interpretations was tracked using retrospective protocols, in which the expert showed that to suggest the interpretations, he would make associations of characteristics visualised in the rock samples. In conclusion, although these features were not fully described in the ontology, there was a relation between visual petrographic features and the suggested interpretation.

In order to extract this causal relation we used the *knowledge graphs* or *k-graphs* first proposed for a different purpose by Leão in [5]. Knowledge graphs were created as a tool for knowledge elicitation in the Cardiology domain problem, specially conceived to represent causal relations between symptoms and diseases. A knowledge graph can be built as a tree, following the specifications below:

- The root node represents an interpretation hypothesis;
- The leaf nodes are the pieces of evidence that indicate some diagnosis;
- The intermediate nodes group those pieces and associate them with the hypothesis.

In order to use k-graphs for knowledge acquisition, the following steps were performed:

- Descriptions of a large number of rock samples were provided by the expert. A rock-sample description contains (i) the petrographic features that can be visualized at the microscope and (ii) the interpretation given by the expert to the rock.

Comentário: Inclua a explicação sobre o novo Kgraph aqui. Verifique cada ocorrência depois.

- A set of possible interpretations to the diagenetic environment of some rock was collected from the descriptions;
- The expert was requested to indicate which petrographic information could support each of the interpretations, and these were then ordered according to their order of significance to the interpretation;
- The hypothesis of interpretation was located in the root of the tree and the pieces of evidence associated with the hypothesis were linked graphically to the root node;
- Significance indexes were attributed to each of the concepts supporting a hypothesis. These expressed the confidence of the expert in the evidence, i.e. the relevance of the evidence in supporting the diagnosis.

Some k-graphs could be elicited, but the eventual gaps in the domain ontology were identified: the rock-sample descriptions did not always exhibit the pieces of evidence that would have supported the interpretation indicated, and, when it did, these items were not the petrographic features described with the ontological vocabulary. The expert used interpreted, expert-level features to support the suggestion of his preferred interpretation rather than the ontological user-level features.

This procedure made evident that the ontology does not include the description of the relevant features. We therefore conclude that there is a significant gap that must be treated between the information that is described and the one that really supports inference (Fig 3(a)). This gap must be handled because a knowledge-based system that provides means drawn from artificial intelligence for suggesting rock interpretations will have user interfaces that should utilize information in the user-level vocabulary.

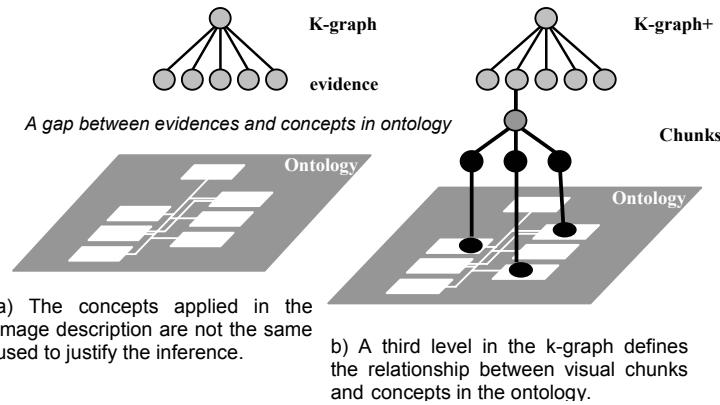


Fig 3. Filling the gap between the information that is described and the information that actually supports inference

As a knowledge acquisition tool, k-graphs are conceived to capture the knowledge at the expert level of reasoning. At this level, the evidential items constructed by experience are more "dense" objects, in the sense that they capture a much greater amount of useful information about of the domain than that a novice can understand and manipulate. For an expert, these items are not complex objects, but units of information that support all the understanding of the domain. We call these units *visual chunks*, in the sense used by VanLehn [14]. Thus, in order to model these visual ob-

jects, and fill the gap between them and the concepts in the ontology, we have devised an enhancement over the "knowledge graphs" scheme of Leão [5], which we have called *knowledge graphs⁺ or k-graphs⁺*.

The k-graphs of Leão were modified to obtain this enhancement by replacing evidence in the model by visual chunks, which are expressed as logical combinations of instances of the concepts in the domain ontology. In image-based reasoning models, a visual chunk has the fundamental role of mapping (or adapting) the expert-level knowledge – the tacit knowledge – to a novice level of knowledge, represented by the simplified (or geometric) domain concepts in the domain ontology. Visual chunks also allow driving the expert-level inference over a problem described by users in novice or intermediate level terms. According to [15], the process of chunking is crucial to the development of expertise. The stimuli that are frequently recognized together tend to assume a peculiar meaning, especially if associated with a particular situation or event. These associated stimuli are used as a cognitive trigger to recall the situation. By doing this, the graphs represent the association between the objects that can be recognized by non-experts users and the objects that support inferences at the expert level.

Thus, in order to elicit the associations of ontological features that formed the chunks, the expert was faced with a number of rock descriptions and was asked to explain *why* a specific solution was chosen instead of another and which domain features supported his decision. The relationship between the pieces of evidence described in the k-graphs and the concepts of the ontology is, thus, filled by intermediate nodes, which describes how those concepts are combined to compose an interpreted visual object (Fig 3(b)). This approach allows us to introduce the visual chunks as a new abstract conceptual type, defined as an aggregation of geometric features described within the domain ontology.

5 K-graphs⁺ as a knowledge-representation formalism

We add to the original knowledge-graph model the idea of "visual chunk" as a new primitive for the representation of visual knowledge. Then, k-graphs⁺ have the role of mapping objects that are described at the user level to the cognitive objects (chunks) applied for inference and, also, expressing how the chunks can support some particular interpretation, producing a knowledge-representation formalism. We detail the characteristics of this representation in [15-18] and present a full example below.

In order to compose the chunk, the domain features were structured in triples concept-attribute-value (CAV) chosen from the shared domain ontology and combined by the logical connectives AND and OR, so as to represent the *domain expressions* abstracted by the expert when recognizing a visual chunk. The AND connective means that every feature is required for the chunk to be recognized, while the OR connective means that any one feature is enough.

Example of k-graph⁺. Fig 4(a) presents the implication relation between the diagenetic environment interpretation of the rock formation (*Continental Meteoric Eodiagenesis Under Wet Climate*) and the visual chunks (e.g., *kaolinization displacive, siderite*). In Fig 4(b), each chunk is decomposed into a group of geological features described in terms drawn from the ontology as concept-attribute-

value (CAV) triples. For example, one feature that a geologist can visualize in a rock sample is that its main *Constituent Name* is *Siderite*. As this information is part of the *Diagenetic Composition* of the rock sample, the whole information is modelled as a feature, that is, a CAV, like "*DiageneticComposition.ConstituentName = Siderite*". The grey nodes represent AND connections and the black nodes are OR connections. In order to recognize the chunk named *Siderite*, all the features and the possible alternative values need to be described by the user.

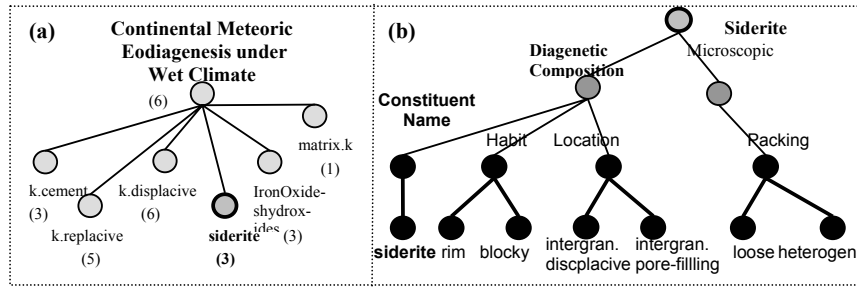


Fig 4. Morphology of knowledge graphs⁺ employed to Petrography

6 Conclusion

The study of expert skills in the geological domain has demonstrated that experts develop a wide variety of representations and hierarchies - which, moreover, differ from what is found in the domain literature. Experts also retain a very large number of symbolic abstractions of images, which we call visual chunks. The chunks hold links with the internal hierarchical arrangement of knowledge, and play an important role in guiding the inference process. These cognitive resources integrate the collection of tacit expert knowledge in geological reservoir characterization.

An association of k-graphs and case analysis has turned out to be an effective tool in helping to make explicit and acquire the declarative knowledge and causal relations of the domain, which were not evident in elicitation sessions conducted on a traditional knowledge-acquisition basis.

The domain model, in this work, expresses knowledge over two levels: the externalization level, which describes the concepts at an intermediate (between novice and expert) stage of expertise, and the expertise level, which has the meaning given above. We have introduced the concept of visual chunk as a primitive for representation. We believe that it should be included in the armoury of any good framework for knowledge acquisition, e.g. the CommonKADS library.

Acknowledgements

This work is supported by the Brazilian Project Financing Agency – FINEP, within the special program CTPETRO.

References

1. Nonaka, I., H. Takeuchi, and H. Takeuchi, *The knowledge-creating company: how Japanese companies create the dynamics of innovation*. 1995, New York: Oxford University Press. xi, 284.
2. Shadbolt, N.R., K. O'Hara, and L. Crow, *The Experimental Evaluation of Knowledge Acquisition Techniques and Methods: History, Problems and New Directions*. International Journal of Human-Computer Studies, 1999. **51**(4): p. 729-755.
3. Gaines, B.R. and M.L.G. Shaw, *Personal Construct Psychology and the Cognitive Revolution*. 2003, University of Calgary - Knowledge Science Institute: Cobble Hill. p. 30.
4. Abel, M., J.M.V. Castilho, and J. Campbell. *Analysis of expertise for implementing geological expert systems*. in *World Conference on Expert Systems*. 1998. Mexico City: Cognizant Communication Offices.
5. Leão, B.F. and A.F. Rocha, *Proposed methodology for knowledge acquisition: a study on congenital heart disease diagnosis*. Methods of Information in Medicine, 1990(29): p. 30-40.
6. Sowa, J.F., *Conceptual structures: information processing in mind and machine*. 1984, Reading: Addison Wesley.
7. Benjamins, V.R. and D. Fensel, *Editorial: problem-solving methods*. International Journal of Human-Computer Studies, 1998. **49**(4): p. 305-313.
8. Schreiber, G., et al., *Knowledge engineering and management - The CommonKADS methodology*. 1999, Cambridge: The MIT Press. 465.
9. Duda, R.O., et al., *Development of the PROSPECTOR consultation system for mineral exploration*. 1978, Stanford Research Institute International: Menlo Park.
10. Schultz, A.W., et al., *XEOD: an expert system for determining clastic depositional environments*, in *Geobyte*. 1988: [S.l]. p. 22-32.
11. Gappa, U. and F. Puppe. *A study of knowledge acquisition - experiences from the SISYPHUS III experimentation for rock classification*. in *Workshop on Knowledge Acquisition, Modeling and Management*. 1998. Alberta, Canada: Voyager Inn.
12. Wagner, W.P., Q.B. Chung, and M.K. Najdawi, *The impact of problem domains and knowledge acquisition techniques: A content analysis of P/OM expert system case studies*. Expert Systems with Applications, 2003. **24**(1 January 2003): p. 79-86.
13. Ericsson, K.A. and J. Smith, *Toward a general theory of expertise: prospects and limits*. 1991, New York: Cambridge University Press.
14. VanLehn, K., *Problem-solving and cognitive skill acquisition*, in *Foundations of Cognitive Science*, M.I. Posner, Editor. 1989, The MIT Press: Cambridge. p. 526-579.
15. Abel, M., et al., *Visual knowledge modelling and related interpretation problem-solving method*. Conferencia Latinoamericana de Informática - CLEI2002, 2002.
16. Silva, L.A.L., et al. *An Image-Based Reasoning Model for Rock Interpretation*. in *WORKSHOP ON INTELLIGENT COMPUTING IN THE PETROLEUM INDUSTRY in the 18th International Joint Conference in Artificial Intelligence, 2003*. 2003. Acapulco - Mexico: Proceedings of the Second Workshop Intelligence Computing in the Petroleum Industry.
17. Abel, M., et al., *PetroGrapher: managing petrographic data and Knowledge using an intelligent database application*. Expert Systems with Applications, 2004.
18. Mastella, L.S., *Analysis of representation formalisms for modeling inferencial knowledge*. (Original in Portuguese: *Análise de Formas de Representação para Modelar Conhecimento Inferencial*), *Research Report*, in *Pos-Graduação em Ciência da Computação*, UFRGS, Porto Alegre, Brasil, 2002.

Comentário: Verificar se monografias são descritas assim. Confiro com o jac.

Comentário: Tirar refe 16 e verificar as demais se estão citadas.