AN INTELLIGENT SYSTEM FOR AUTOMATED CLASSIFICATION DRILLING REPORT

Clarice Augusta Carvalho Rabelo
UNICAMP/FEM/DEP
C.P. 6052, Campinas, São Paulo, Brasil
clarice@dep.fem.unicamp.br

Ivan Rizzo Guilherme
UNESP/IGCE/DEMAC - Rio Claro
ivan@rc.unesp.br

José Ricardo Pelaquim Mendes
UNICAMP/FEM/DEP
C.P. 6052, Campinas, São Paulo, Brasil
jricardo@dep.fem.unicamp.br

Celso Kazuyuki Morooka
UNICAMP/FEM/DEP
C.P. 6052, Campinas, São Paulo, Brasil
morooka@dep.fem.unicamp.br

João Carlos Ribeiro Plácido
PETROBRAS/CENPES/PDP/TEP
Cidade Universitária, Quadra 7, Rio de Janeiro, Brasil
jcrp@petrobras.com.br

Abstract. *Petroleum wells drilling activity is complex and extremely dependent on technical skills and engineering competence. During drilling process, a great amount of information is generated, representing an important source of knowledge. Part of this information is automatically obtained from equipments, while other are reported, generally, in text format. Such information is analyzed by engineers involved with well design projects, evaluating operation executions and identifying problems to provide solutions. In this way, the technology evolution in information system area has provided greater capability to generate and store data in text format, which are an important source of institutional knowledge of oil companies. This paper presents an application of an intelligent system for automated text analysis using a sample of drilling records to demonstrate how the methodology works, which can be used in different areas of petroleum and mechanical engineering. The methodology used here consisted basically on organizing the knowledge related to drilling process by elaborating a dictionary of some typical problems. The whole process was assisted by a drilling expert, also using a software to collect the knowledge from the texts. Finally, a sample of drilling reports was used to test the system, evaluating its performance on automated text classification.*

Keywords*: Drilling problems, Intelligent System, Text Processing*

1. Introduction

Drilling activity in oil industry is complex and it involves great amount of investment. In this way, one of the most important goals pursued by the sector is to guarantee that drilling operations are carried out in a safe and economic manner. However, that is not an easy task and it requires a specialized knowledge, technical skills and engineering competence in order to achieve these goals.

According to Plumb *et al.* (1999), great part of the time and the cost spent drilling is related to the process of getting into the reservoir. Many problems might occur during drilling operations and the imperative aim in this practice should be the control of all risks involved with the activity.

The complexity and the search for more efficient process have driven the sector to a great level of specialization (Bourgoyne *et al.*, 1986). According to Mendes *et al.* (2003), such development has caused increasing sophistication in drilling operations, requiring greater amount of information to better plan and design wells.

In this context, most of petroleum companies have developed a method to generate and store great part of information obtained during drilling process. It is very common to make a record of all operation details, including process performance and problems occurred. These records provide to the company a large amount of data that can be analyzed and used in future projects in order to optimize drilling activity (Edwards *et al.*, 2001).

Based on this analysis, significant decisions and actions can be carried out. Particularly in well designing and planning, this process of analyzing previous well drilling records can be very useful to support decisions. Thus, records in text format such as drilling operation technical reports are very important to learn from successful events and abnormal occurrences and to use this experience in a new similar well construction (Miura *et al.*, 2003).

However, processing this amount of information is not simple. Manual classification and consolidation of these reports in the past were very time-consuming, representing a six-month job of an expert to classify and consolidate half-year period reports. But since technology in text processing has developed during the past years, today is possible to handle and analyze huge amount of data using automated systems.

This paper presents a methodology based on automated text analysis that represents a very useful tool to handle great amount of text information. As an example of its application, the system was used to analyze a sample of drilling records provided by petroleum companies. A software was used to create a drilling problem dictionary based on the records available with the assistance of a drilling expert. As result, a report containing all problems identified during different drilling operations was generated. The system performance was also evaluated by making a comparison between the automated and the manual analysis previously made.

It is relevant to say that the aim of this paper is to show how the methodology presented here can be useful to process information not only in drilling operations, but also in different areas of petroleum and mechanical engineering.

2. Methodology for Automated Text Processing

The methodology for automated text analysis is an interactive process between the expert and the software tool based on intelligent text processing. Its application allows a specialist (domain users) to improve his expertise and to learn more about the domain analyzed (Rocha *et al.*, 1992).

According to Miura *et al.* (2003), text analysis is, generally, a very hard task due to the necessity of having a complete dictionary with a large number of terms available and the semantic ambiguity of the language used. However, this analysis can be easier in more specialized texts, when the number of words and the semantic used in their internal communications are sharply defined in the domain of the specific community.

The methodology presented in this paper is an important tool that can be used to process large amount of information. Differently from usual methodologies available in literature, in text processing used here, texts files are not only characterized and classified at word level, but also in phrase level.

The whole process is based on the creation of a knowledge structure represented by a phrase dictionary. This dictionary is actually a set of word combinations with specific meaning for the context considered. In this paper, the phrase dictionary generated represents a structured knowledge of drilling problems and every phrase that it contains represents a different abnormality related to drilling operations.

The phrases from the dictionary can be generated with different number of words combined. All of them must contain at least one word with stronger meaning that can be complemented by other word(s) in a way that the phrase correctly corresponds to a knowledge description. This process of associating the words is defined by the grammar.

The phrase dictionary is created by using the text analysis tool. A sample of texts from the domain analyzed is required to generate a list of words with specific meaning for the context of study. The word grammar is defined with the assistance of an expert, generating the word dictionary. The phrase dictionary is then created by selecting a set of words combinations that represent the knowledge in the domain. Summarizing the applied methodology:

1. *(Expert)* Selection of a sample of texts from the domain analyzed;
2. *(Text Analysis Tool)* List of words extracted from the texts;
3. *(Expert)* Selection of words with strong meaning for the context considered and grammar definition. This is the final step from the process of creating the word dictionary;
4. *(Text Analysis Tool)* Text processing to create a list of word combinations based on the grammar definition;
5. *(Expert)* Selection of phrases from the list of word combinations with strong meaning for the context, generating the phrase dictionary. It is also possible to refine the grammar or manually adjust word combinations, for example, by adding restriction terms. In this case, if the dictionary is used to analyze a text that contains the phrase and the restriction term in the same event description, the system will not be able to identify it;
6. *(Text Analysis Tool)* Using the phrase dictionary, the text analysis tool can be used to process all new desired set of texts and to classify them under any kind of abnormality. As result, a map of abnormalities and operations within the interest of the expert can be obtained.

To better understand the process described above, a scheme from the methodology for automated text analysis is shown in Fig. 1.

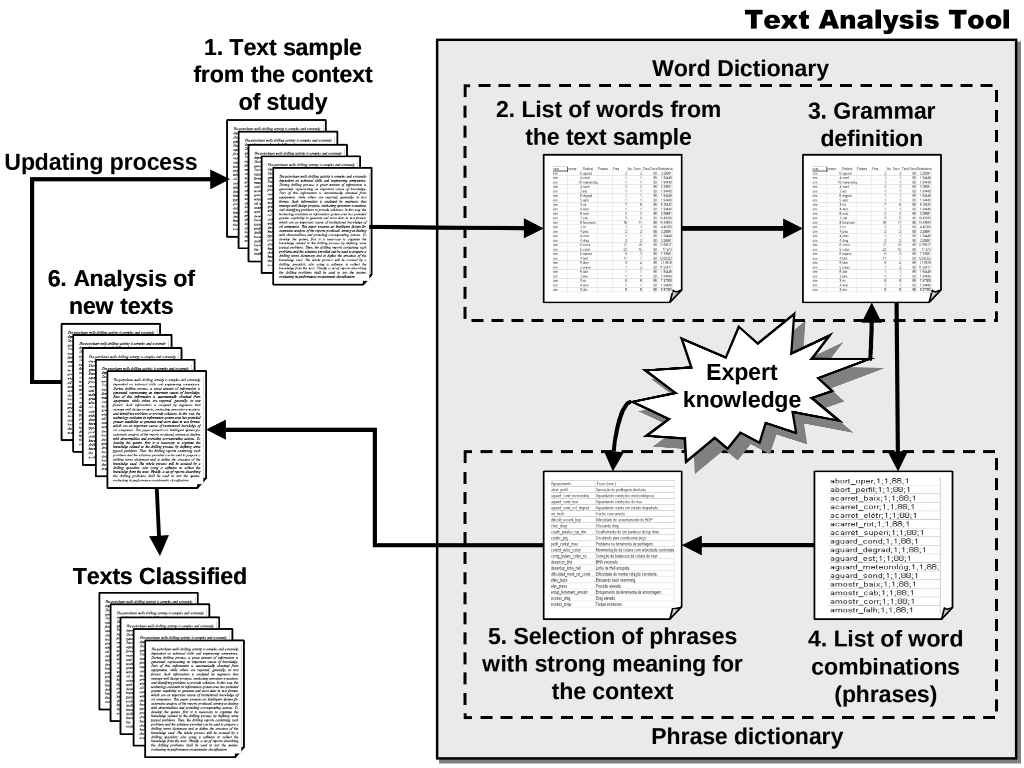

Figure 1. Methodology for automated text processing

By using this methodology, abnormal occurrences can be automatically classified and consolidated based on the experience of previous drilling operations. Its application might reduce the job effort from the regular six-month manual man job of an expert to one-week interactive man-computer job (Miura *et al.*, 2003).

Another important aspect to be observed in this process is the evolution of well engineering knowledge used in this activity. The introduction of new technologies or new concepts in the area may require an update of the phrase dictionary due to its gradual loss of performance verified through the years. However, this updating plan does not require either large amount of investment nor professional function deviations. It is recommended to update the system after each year (Morooka *et al*., 1993).

3. An Application of Automated Text Processing to Generate a Dictionary of Drilling Problems

The methodology presented in this paper is applied to a community involving personnel in well construction activities in petroleum upstream industry and they refer to drilling and completion operations. As already mentioned, many petroleum companies generate daily operational reports containing descriptions of all occurrences detected for, at least, at a half-hour time basis. These records represent a great source of knowledge since they contain all events, classified by operations, in a time sequence, including abnormal occurrences detected while drilling. The importance of such knowledge was already discussed in the introduction and it is related to the learning process that might help the companies to manage the risks involved in the activity.

Based on these reports, the methodology presented in this paper is used to create a knowledge representation (phrase dictionary) of drilling problems and use this knowledge to build an intelligent system to analyze and classify a sample of reports that are in text format.

Previously to the text analysis, some major drilling problems were studied by making a research in the literature available. This procedure was necessary to better understand some drilling abnormalities and how they can be detected during operations.

The text processing was completed by running two different analysis: a manual and an automated one. A sample of fifteen drilling records was used, each record associated to different operations, including directional, horizontal and vertical wells. Considering all records together, there were more than four thousands sentences or activities descriptions to be analyzed.

The manual analysis consisted basically on reading the reports, identifying abnormal occurrences and actions taken to mitigate the potential injuries and mapping them. The manual processing provided a better comprehension of the natural language used on the reports to describe the problems.

Although this manual analysis seems to be a simple procedure, the process of reading all texts is very time consuming and also exhausting. The large number of different events described also makes the analysis much more susceptible to errors.

The automated text processing was initiated based on the same sample of drilling records used for the manual analysis. First, the automated text analysis tool was used to generate a word dictionary based on the sample of texts given. Some examples of words identified are shown in Tab. 1. The software also allows the manual addition of new words.

Table 1. Word dictionary created using Text Analysis Tool.

Use	Radical	Suffix	Suffix Freq.
Yes	drill	ing	4
Yes	fluid		
Yes	loss	es	9
Yes	mud		
Yes	formation		

In the example given in Tab. 1, the words "mud" and "formation" were added manually. This manual procedure should be done when it is considered that might have different ways to describe the same abnormality. For example, the problem of "Loss of drilling fluid" can be also identified as "Loss to the formation" or "Mud loss".

At this point, it is important to make a good grammar definition. The reason is that the grammar drives the process of associating the words, creating a sentence that represents the problem description. For the same example given previously, by properly defining the grammar, four different word combinations were created (Tab. 2).

Table 2. Example of phrase dictionary associated to the problem of "Loss of drilling fluid".

Word combination	Restriction	Phrase (Problem Description)
drill_fluid_loss	without	Loss of drilling fluid
fluid_loss	without	Loss of drilling fluid
mud_loss	without	Loss of drilling fluid
format_loss	without	Loss of drilling fluid

Some of the phrases were also added manually based on the same assumption presented to the manual addition of new terms to the word dictionary. The initial research about drilling problems and the manual analysis were very important to this process of elaborating new word combinations, considering different abnormality descriptions as possible.

Processing all records available, also considering the combinations added manually, generated a large number of associations to define different abnormal occurrences. However, some of them had no real meaning for the context treated here and only a few of them were used to generate a dictionary of drilling problems.

Another important adjustment that can be made is the addition of some negation terms that are used to create new combinations with opposite meaning. Some words can be also used as restriction during the problem identification process. In this paper, for example, the word "without" was used as a restriction to the phrases in Tab. 2. So, when the system was used to analyze a new sample of texts, events descriptions that contained those phrases and the word "without" were not identified.

It is also important to remember that this is an interactive process, with the possibility of modifying and adjusting the system according to the convenience. More associations can be created later, updating the system depending on the necessity.

The final step consisted on analyzing a new sample of reports using the phrase dictionary elaborated. The results obtained with the system application are shown in the next section.

4. Results of Report Analysis

The manual analysis provided the identification of one hundred and twenty four different types of abnormal occurrences in the sample of texts analyzed. The same problems were considered in the automated text processing, resulting on the creation of almost one hundred and fifty word combinations to represent the abnormalities.

As result from manual analysis, it was generated a report containing all problems identified, depth on which they have occurred, date and hour. An example of such report is shown in Tab. 3.

Table 3. Result of manual text analysis.

Date	Initial hour	Final hour	Duration	Initial depth	Final depth	Problem description
06/08/03	13:00	14:30	1.5	0	0	Drillstring motion with controlled speed
08/08/03	14:00	17:30	3.5	1121	1121	Leaking problems with housing
09/08/03	8:00	10:30	2.5	1121	1121	Assembly equipment problems
11/08/03	20:00	20:30	0.5	1536	1536	Bad weather conditions
13/08/03	9:00	14:00	5	2350	1117	Drillstring motion with controlled speed due to resistant points
16/08/03	11:00	14:00	3	1058	1058	Failure in rig functions tests
21/08/03	7:00	20:00	13	2415	2415	Geopilot orientation problems
22/08/03	7:30	14:30	7	2340	2415	Drillstring sticking
22/08/03	20:00	0:30	4.5	0	0	Problems with LWD tool
22/08/03	15:00	22:00	31	2415	0	Geopilot orientation problems
25/08/03	21:30	23:30	2	2704	2704	Top drive repair
26/08/03	2:30	3:00	0.5	2930	2930	Loss of drilling fluid
29/08/03	15:30	13:30	118	3286	3286	Drillstring sticking

A similar result was obtained from the automated text processing. The report contained information like the problem description, corresponding word combination (problem representation), frequency, relevance and number of documents on which they were identified. Table 4 contains an example of the report generated.

Table 4. Example of report generated containing different problems descriptions used on the analysis

Word combination	Restriction	Phrase (Problem description)
abort_log		Logging operation aborted
wait_weath		Waiting on weather
check_drag		Drag checking
condition_well		Well conditioning
clean_hole		Poor hole cleaning
loss_form	without	Drilling fluid loss
drop_press		Pressure dropping
stick_drillstr		Drillstring sticking

The main difference between both analysis was the period of time required to conclude the process. The manual analysis was concluded in two weeks, including time spent to generate the final report, while the intelligent system applied was almost fifty per cent less time consuming. It is relevant to say that most of the time spent with the automated processing was related to the phrase dictionary elaboration and some adjustments, which means that, when this step is concluded, the real time spent to classify new samples of texts is only a few minutes.

A second difference between the manual and the automated analysis was related to a few occurrences that were identified in one method, but not in the other one. This observation shows that both methods have failures, but the automated system can be pointed as more trustable since a manual analysis of very extensive texts as the drilling reports is much more susceptible to human failure.

Another result of the text processing was the classification of the abnormalities identified. The main drilling problems identified were classified in five different areas:

- Equipment problems;
- Poor hole cleaning;
- Drillstring sticking;
- Bad weather conditions;
- Drilling fluid losses.

The most common abnormalities were associated to equipment problems. Most of these abnormal occurrences identified did not contain the problem description itself, but usually only a mention about the type of equipment and the action taken to treat the problem, like replacement or repair. The equipment that had more occurrences in the reports analyzed was the top drive.

Problem with hole cleaning was also very common in the reports of all wells studied. Other abnormalities that were also identified, but with smaller frequencies, were drillstring sticking, bad weather conditions and drilling fluid loss.

By using this problem classification and information contained in Tab. 4, it was generated a pareto graph of abnormalities. Figure 2 presents the results obtained from automated text analysis.

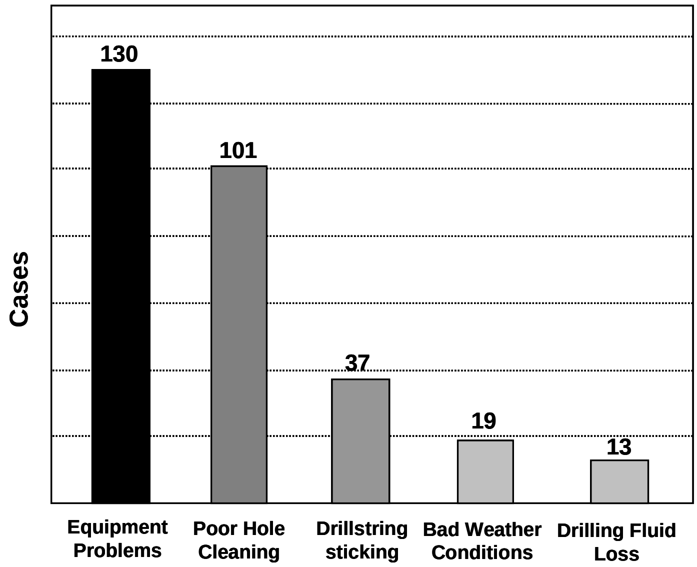

Figure 2. Pareto graph of abnormality (information obtained from different wells drilled).

This graph can be used to illustrate the most frequent abnormalities that may occur during drilling operations. This information can be very useful since they show which are the major problems, allowing the companies to prioritize and treat more efficiently the most common ones in order to achieve cost-effective benefits.

It is important to note that the main concern of the methodology applied here is not a pure drilling problem classification. The most relevant purpose of this system is to increase the operation efficiency by decreasing the loss of time and money with the application of a treatment procedure to prevent problem recurrence.

However, the system presented in this paper requires a regular procedure of updating. This procedure is necessary because there is always the possibility of different problems descriptions be used in drilling reports. In this way, this updating process is recommended to guarantee a greater efficiency on the identification of abnormal occurrences.

In similar study to this one made by Miura *et al.* (2003) to completion operations, the efficiency obtained with the dictionary elaborated from half-year period dropped from 80% (first years) to 55% after three years processing texts. One of the main causes pointed in his work that might contribute to this observation is that staff member in charge of figuring out the problems (analysts) are concerned with only a few concepts (major problems). But, even with this loss of performance, the methodology provides a minimum efficiency of more than 50% percent, which is more than a manual analysis can provide.

5. Conclusion

Drilling operations complexity has grown due to the also increasing extreme conditions, with deeper wells drilled in higher temperatures and pressure reservoirs. In this context, the amount of information required to design wells is getting greater, forcing the companies to deal with larger quantities of information, storing and analyzing them in order to drill further wells more efficiently.

This paper presented an intelligent system used to process the information generated during drilling operations. The results obtained by applying this methodology have shown that the system can be very useful as an automated procedure to identify and classify abnormal occurrences in well engineering activities. The importance of such process was discussed in the introduction of this paper.

The automated classification is not the only result that can be obtained by using the system. Information about the problems frequencies can also be used to identify the major problems, allowing the companies to prioritize the most common ones. Further, the knowledge generated here can be very useful to training purposes and operation activities planning in petroleum engineering.

As next steps of the present study, new text processing techniques are under research. This paper represents an initial step to the efforts that have been made to propose a system capable of dealing automatically with the data generated in real-time during drilling operations and use them to anticipate the possibility of having problems.

6. References

Bourgoyne, A.T., Millheim, K.K., Chenevert, M.E. and Young, F.S., 1986, "Applied drilling engineering", SPE Textbook Series, vol. 2, USA, 502p.

Edwards, T. M., Murphy, J. F., Goraya, S., Harrold, T., Holt, J., Lechner, J., Nicholson, H., Standifird, W. and Wright, B., 2001, "Avoiding Drilling Problems", Oilfield Review, Summer Edition, pp 33-51.

Mendes, J.R.P., Morooka, C.K. and Guilherme, I.R. 2003, "Cased-based reasoning in offshore well design", Journal of Petroleum Science and Engineering, Vol. 40, pp. 47-60.

Miura, K., Guilherme, I.R., Morooka, C.K. and Mendes, J.R.P, 2003, "Processing Technical Daily Reports in Offshore Petroleum Engineering – An Experience", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 7, No. 2, 223-228p.

Morooka, C.K., Rocha, A.F., Miura, K. and Alegre, L., 1993, "Offshore Well Completion Operational Knowledge Acquisition and Structuring", Eleventh International Offshore Mechanics and Arctic Engineering Conference (OMAE), ASME, Glasgow, Scotland.

Plumb, W.A., Gholkar, I.A., Minton, L.C., Fuller, J., Goraya, S. and Tucker, D., 1999, "Managing Drilling Risk", Oilfield Review, Summer Edition, pp 2-19.

Rocha, A. F., Guilherme, I. R., Theoto, M., Miyadahira, A. M. K. and Koizumi, M. S., 1992, "A neural Network for extracting Knowledge from Natural Language Data Bases", IEEE Transactions on Neural Network, Vol. 3, No. 5.