INTELIE by Viasat™

Extracting Events From Daily Drilling Reports Using Fuzzy String Matching

Mariana Oliveira, Adriano Mourthe and Maria Clara Duque

May 19, 2020

Viasat™

Agenda

01 **Introduction**

02 **Methodology**

03 **Results and Discussion**

04 **Conclusions**

Introduction

- Daily Drilling Reports (DDRs):
  - Standard practice across companies;
  - Describe actions and events during well construction stages.

Rich Source of Information

From previous constructed well reports:

1. Good Practices
2. Intervention Operations
3. NPT events
4. **ILT events**

ILT

Invisible Lost Time

Difference between **actual operation** duration and a **best practice** target.

Introduction

- DDR historical data: Extract information to drill a new well more efficiently.

Historical well drilling data **search**

Drill a new well more **efficiently**

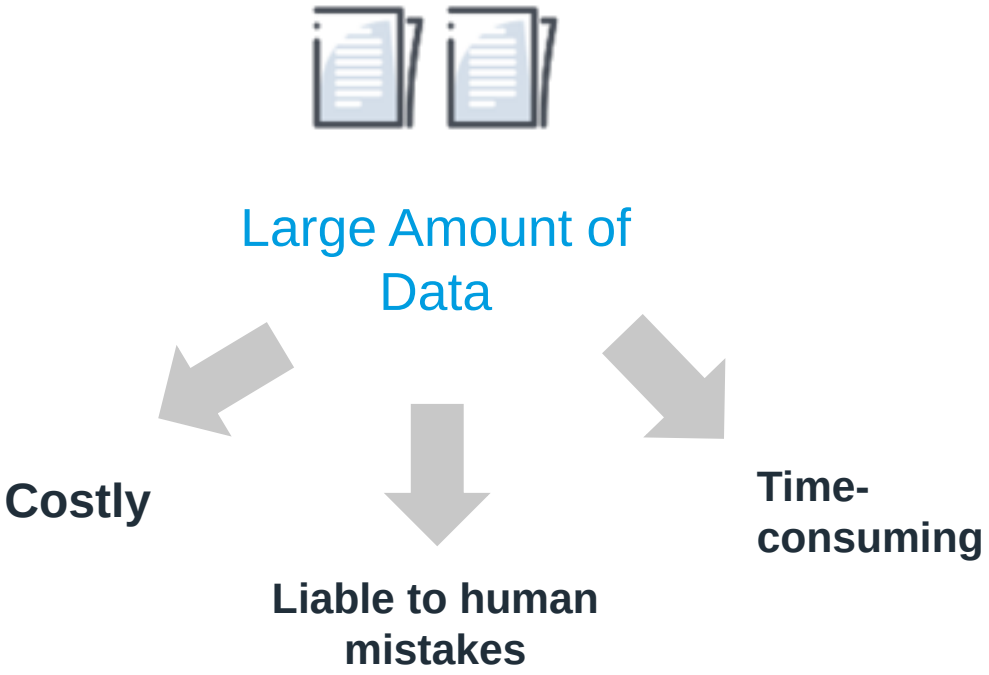
Historical data

Identify **ILT events**:

- Contribute to higher operational efficiency.
- Cost Reduction of drilling operations.

Introduction

- Manually analyzing DDR text reports is a difficult task



Alternative Approach:

- **NLP techniques** to automatically process text data and extract information from DDRs texts.

Introduction

Objectives

- Automatically extract possible **Invisible Lost Time (ILT)** events from DDRs that may indicate risks or low operational efficiency.
- Propose an NLP methodology to automate information retrieval and assist the decision-making and risk prevention process.

Methodology

Rule-based Language Processing

Methodology

Rule-based Language Processing

Reduce Noise:

- Lowercase conversion
- Symbols removal
- Blank spaces removal

Methodology

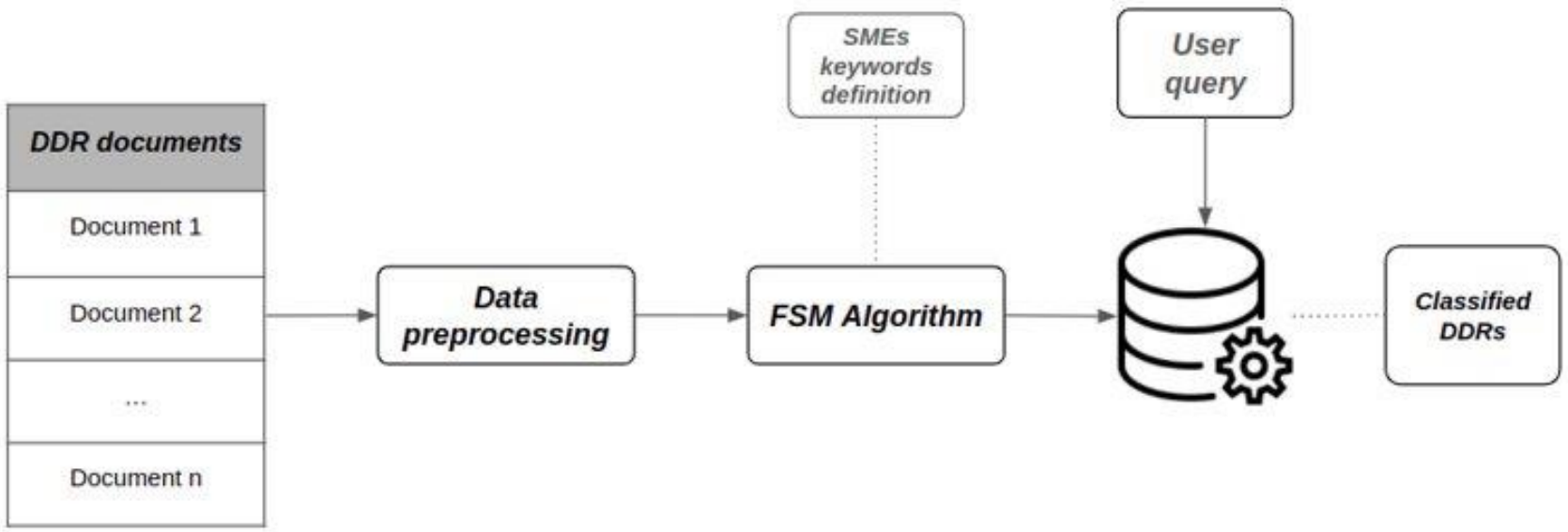
Rule-based Language Processing

Search for pre-defined expressions and keywords:

- Allows limited number of error
- Find matches with misspelled words and different prefixes or suffixes.

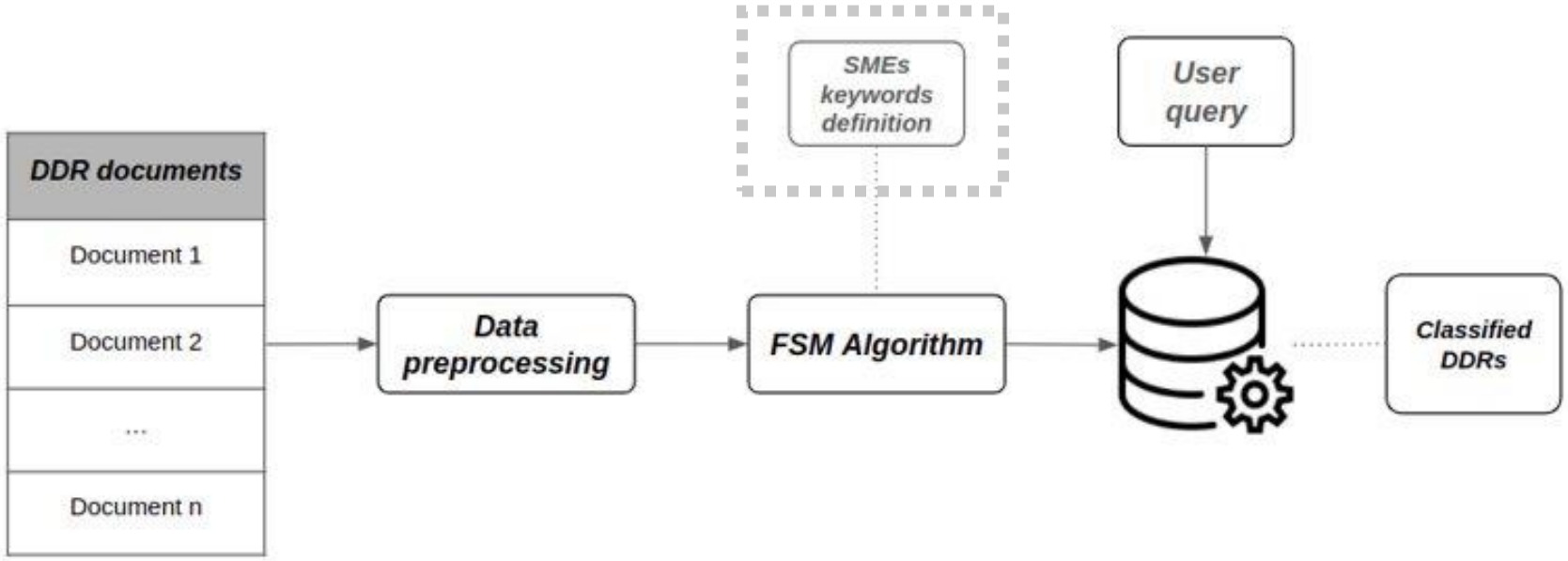
Methodology

Rule-based Language Processing



Methodology

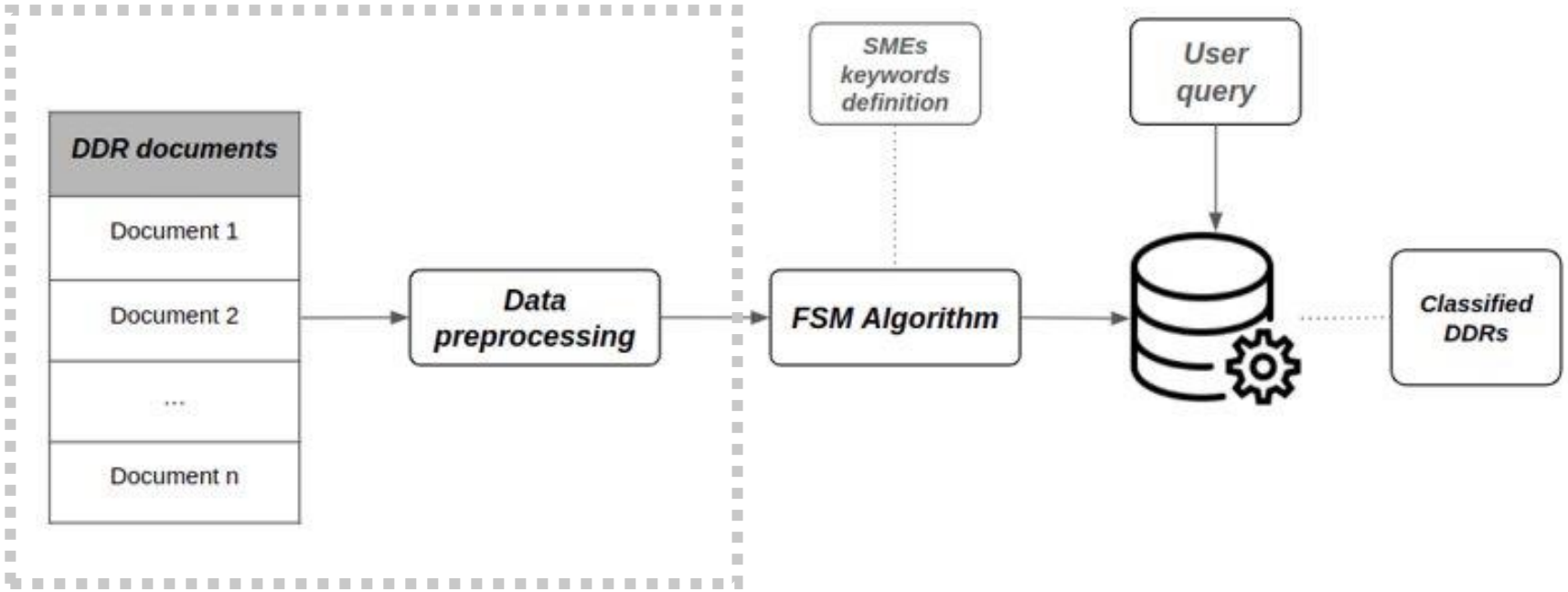
Rule-based Language Processing



- SME read DDR text samples
- Define keywords related with 6 ILT problems:
  - Vibration Issues,
  - Circulation Loss,
  - Equipment Failure,
  - Geological Risk,
  - Hole Conditioning Issues,
  - Directional Drilling Issues.

Methodology

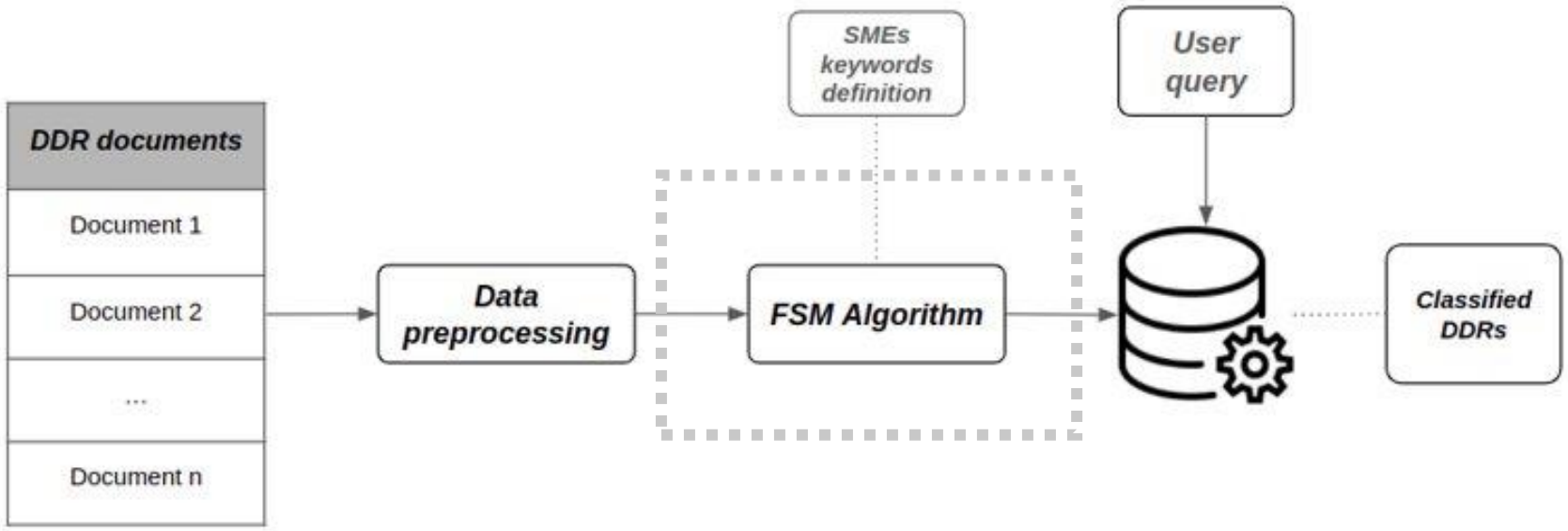
Rule-based Language Processing



- Raw documents are preprocessed to reduce noise.

Methodology

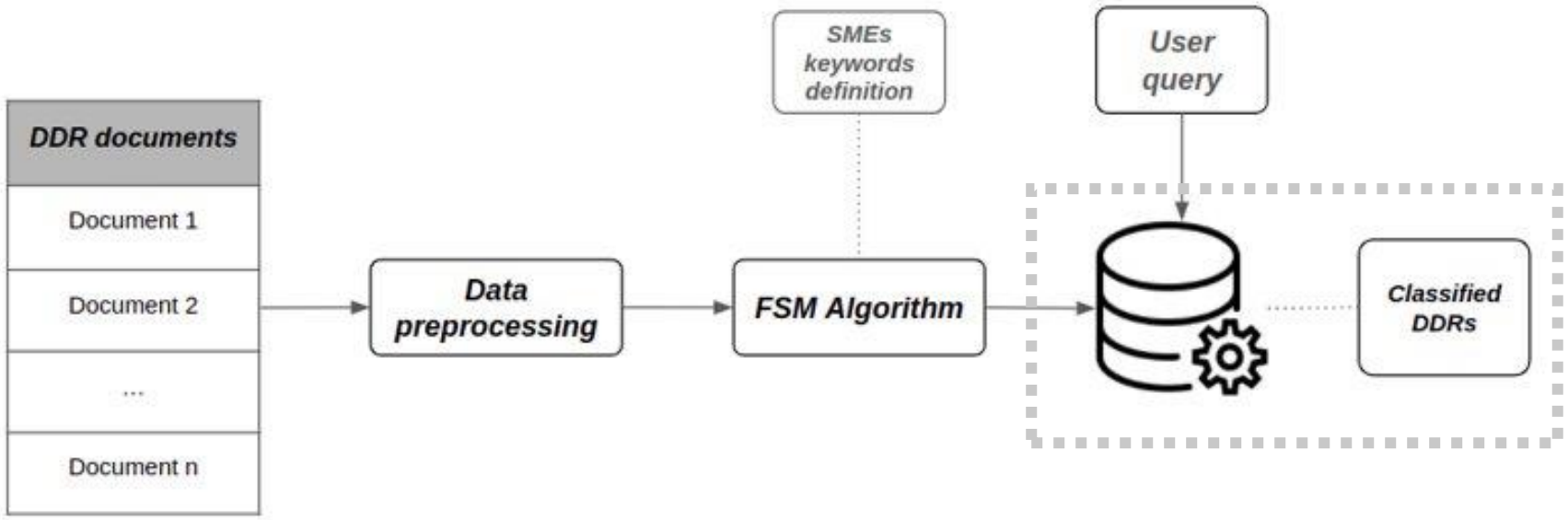
Rule-based Language Processing



- **FSM algorithm** is applied, using previously annotated expressions.
- **FSM**: quantifies dissimilarity between two strings.
  - Distance: Normalized Damerau-Levenshtein (NDL).
  - Match:
    - distance ≤ 30%.

Methodology

Rule-based Language Processing



- When the algorithm finds an **approximated expression**, the record is classified.

Methodology

Rule-based Language Processing

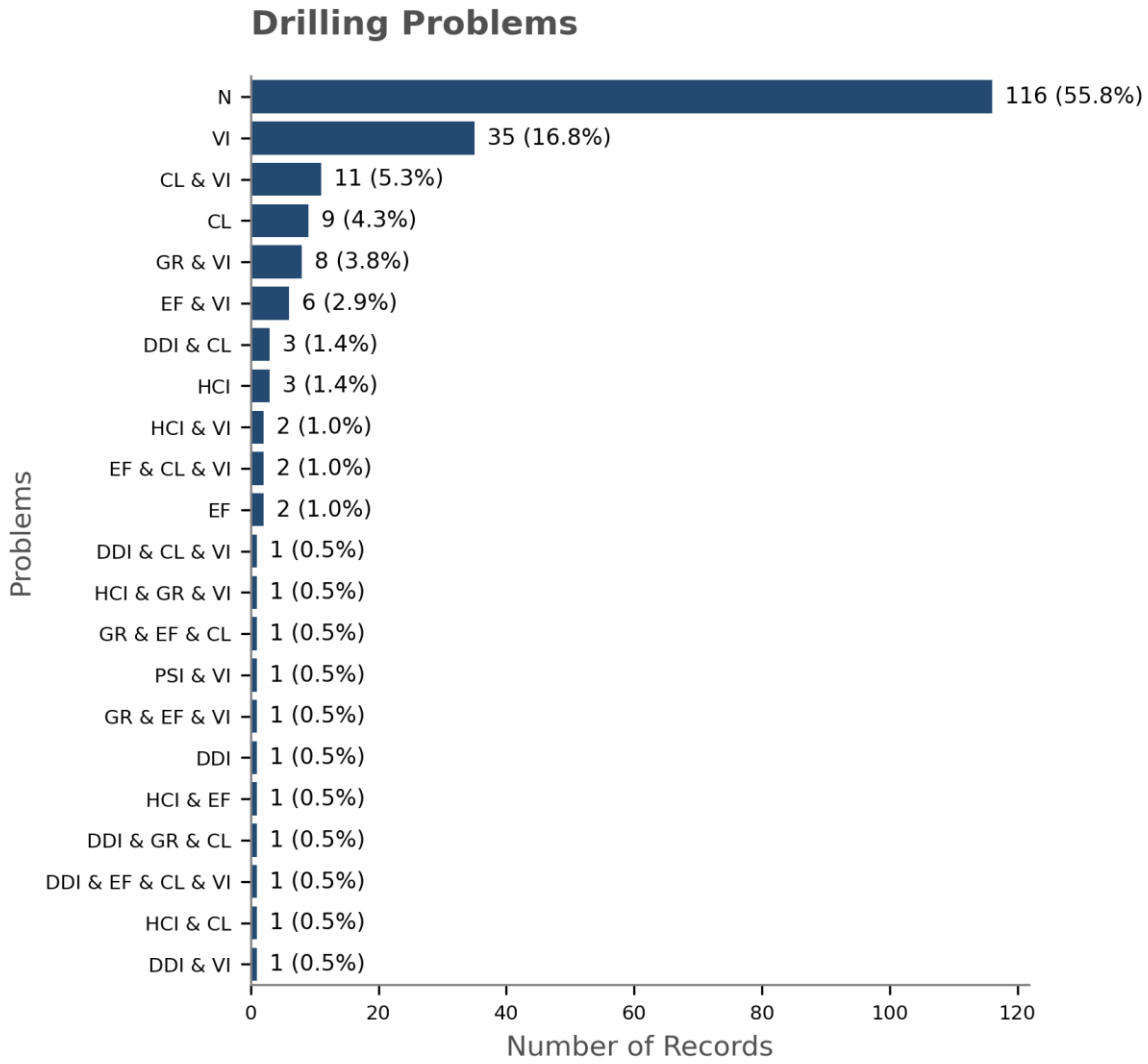
- Example:
  - Annotated Expression: erratic Torque
  - ILT Problem: Vibration Issues

Original Record	Preprocessed Record	Record with Expressions
Control ROP 75-115 fph due to erraatic torque.	*control rop fph due to erraatic torque*	*control rop fph due to erraatic torque* Match: erratic torque **Label: Vibration Issue**

Results and Discussion

- Dataset:
  - 208 Real-world DDRs records from a drilling company
  - ILT keywords and expressions mapped from SME

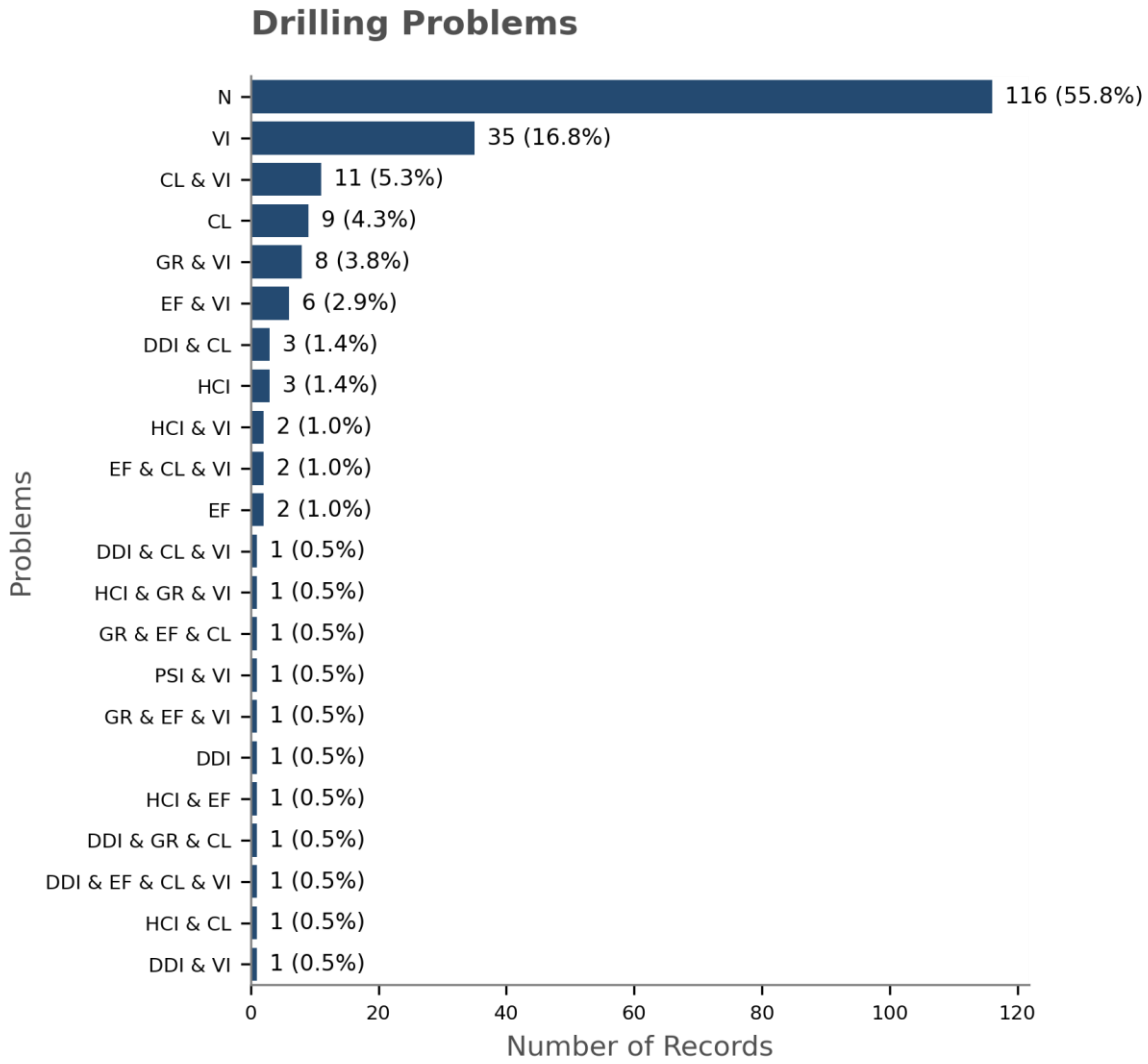
- Experimental Results:
  - Normal Operations: 208 DDRs (~2600 hours)
    - 92 possible ILT descriptions (~900 hours)



N: Normal; VI: Vibration Issues; CL: Circulation Loss; GR: Geological Risk; EF: Equipment Failure; DDI: Directional Drilling Issues; HCI: Hole Conditioning Issues.

Results and Discussion

- Experimental Results:
  - Multilabel task:
    - Multiple labels may be assigned to each text.
    - One single DDR record can have different ILT problems.



N: Normal; VI: Vibration Issues; CL: Circulation Loss; GR: Geological Risk; EF: Equipment Failure; DDI: Directional Drilling Issues; HCI: Hole Conditioning Issues.

Conclusions

- This paper applied FSM methodology that combines Specialist knowledge with NLP techniques to find ILT expressions with possible typos and different suffixes or prefixes.
- The methodology was tested at real-world DDRs. Considering normal drilling operations, almost **35%** of the hours were identified as ILT.
- The methodology was able to identify ILT problems even with the great quantity of misspelling along the text, facilitating the search for problematic DDRs, reducing time and automating user workflow.

Thank you!