DEVELOPING NLP TOOLS FOR SHARING OF INDIGENOUS & COMMUNITY KNOWLEDGE

Alex Chow
Huaiwen Dong
Yingqiu Kuang

NATIONAL ENERGY BOARD | UBC DATA SCIENCE INSTITUTE
DATA SCIENCE FOR SOCIAL GOOD 2019

Final Presentation | 22nd August 2019

OUTLINE

- Project Context
  - Partner Organization
  - Overview of Dataset
  - Project Goals
- Project Overview
  - Proposed Solution
  - Demonstration

PROJECT CONTEXT

Partner organization

Overview of dataset

Project goals

PARTNER ORGANIZATION

National Energy Board

- Regulates pipelines, power lines, energy development, and trade in the Canadian public interest
- Main goal is to minimize harm by considering economic, environmental, and social factors
- Court hearings help inform decision-making process

OVERVIEW OF DATASET

Transcripts from NEB Court Hearings

8554. **MS. HUME:** Emma Hume, legal counsel for Blueberry River First Nations.

8555. **MS. GAGNÉ:** Mr. Chair, I'm Carol Gagné, Board counsel.

8556. I understand that we also have Michael Sawyer. He's not in the room currently, but will be here today. Registered electronically we have Kenji Tokawa and Naomi Owens for Saulteau First Nations. Also appearing later today will be Anne Muter from West Moberly First Nations. And Ms. Allisun Rana was here yesterday representing Doig River First Nations. As well as Kate Gunn representing Wilp Luutkudziiwus.

8557. **THE CHAIRMAN:** Thank you.

8558. Any other counsel? I wouldn't think so. Good. Okay.

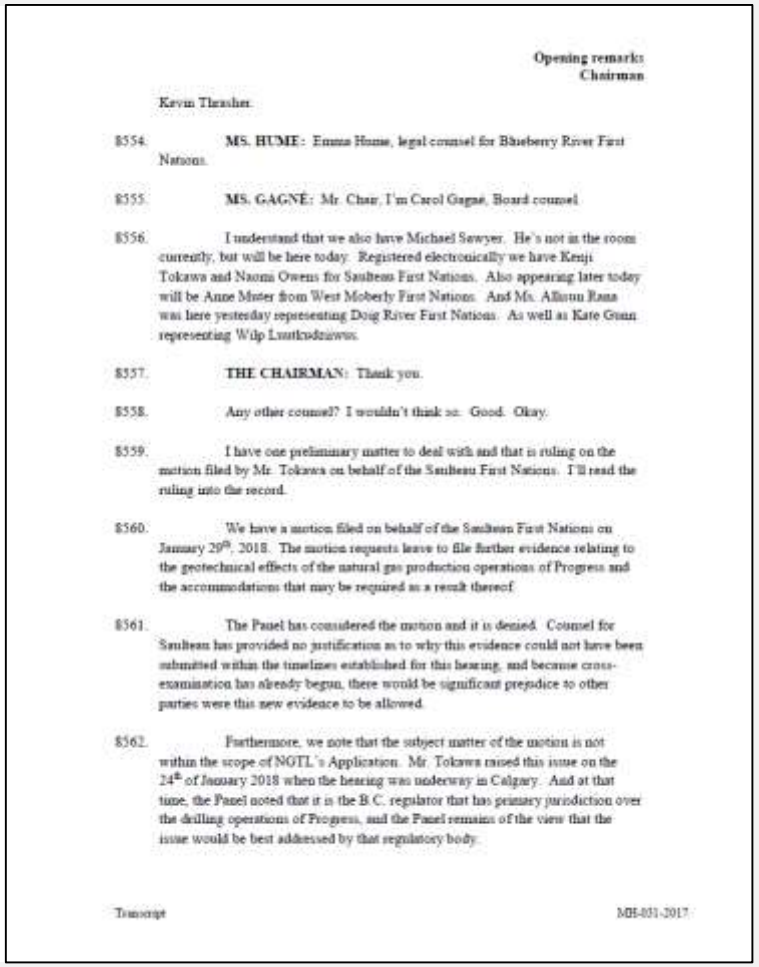
8559. I have one preliminary matter to deal with and that is ruling on the motion filed by Mr. Tokawa on behalf of the Saulteau First Nations. I'll read the ruling into the record.

- Type of hearing
- Location of hearing
- Testimonies
- Support/Concerns

- 60 years of transcript data (1959 – 2018)
- ~ 6000 transcripts in total
- ~ 30 – 200 pages per transcript

PROJECT GOALS

1. Over the last 60 years, who are the active participants in hearings?
2. What are the major concerns and interests in these hearings?
3. What are their attitudes towards these issues?
4. How can we preserve indigenous stories/testimonies from these hearings?



Kevin Thrasher.

8554. **MS. HUME:** Emma Hume, legal counsel for Blueberry River First Nations.

8555. **MS. GAGNÉ:** Mr. Chair, I'm Carol Gagné, Board counsel.

8556. I understand that we also have Michael Sawyer. He's not in the room currently, but will be here today. Registered electronically we have Kenji Tokawa and Naomi Owens for Saulteau First Nations. Also appearing later today will be Anne Muter from West Moberly First Nations. And Ms. Allison Ross was here yesterday representing Doig River First Nation. As well as Kate Gunn representing Wilp Luutkudziwus.

8557. **THE CHAIRMAN:** Thank you.

8558. Any other counsel? I wouldn't think so. Good. Okay.

8559. I have one preliminary matter to deal with and that is ruling on the motion filed by Mr. Tokawa on behalf of the Saulteau First Nations. I'll read the ruling into the record.

8560. We have a motion filed on behalf of the Saulteau First Nations on January 29th, 2018. The motion requests leave to file further evidence relating to the geotechnical effects of the natural gas production operations of Progress and the accommodations that may be required as a result thereof.

8561. The Panel has considered the motion and it is denied. Counsel for Saulteau has provided no justification as to why this evidence could not have been submitted within the timelines established for this hearing, and because cross-examination has already begun, there would be significant prejudice to other parties were this new evidence to be allowed.

8562. Furthermore, we note that the subject matter of the motion is not within the scope of NGTL's Application. Mr. Tokawa raised this issue on the 24th of January 2018 when the hearing was underway in Calgary. And at that time, the Panel noted that it is the B.C. regulator that has primary jurisdiction over the drilling operations of Progress, and the Panel remains of the view that the issue would be best addressed by that regulatory body.

PROJECT OVERVIEW

NLP tools

Proposed solution

Demonstration

PROPOSED SOLUTION

1. Extract relevant data (e.g. transcript text) and metadata (e.g. speaker and organization names) from files to create searchable database

2. Use Natural Language Processing (NLP) tools to model topics, analyze sentiment, and identify important sentences/stories

NATURAL LANGUAGE PROCESSING TOOLS

Speaker Recognition & Topic Modelling

	--- EXAMINATION BY/INTERROGATOIRE PAR MS. HUME:		
Speaker 1	8611 **MS. HUME:** Good morning, panel.	Topic 1: Admin	
	8612 NGTL has applied for eight new meter receipt stations under Section 58 of the *NEB Act*, which are identified in the variance application. And when I say new metre receipt stations, you'll understand that those are the meter stations that I'm speaking about. I'd ask that you please just describe, generally, the function of a meter receipt station, what determines its size, and describe, you know, what it looks like and its general appearance.	Topic 2: Meter Receipt Stations	
Speaker 2	8613 **MR. PLATT:** So the receipt meter stations are located within our pipeline corridor. The purpose of the meter stations are to receive gas from our customers who are producing gas in the area onto the North Montney Mainline. They consist of some small diameter piping. They consist of some -- a separator skid, a metering skid and controls, and measurement devices so that we can record in volume of, and the energy of, and the gas quality of the gas being received into the North Montney Mainline system.	Topic 2: Meter Receipt Stations	
Speaker 3	8614 **THE CHAIRMAN:** Mr. Platt, I wonder if you could bring the microphone closer to you when you speak?	Topic 1: Admin	
	8615 Thank you.		
Speaker 2	8616 **MR. PLATT:** So, generally, they would consist of a number of skids. They would consist of -- and depending on the size of the meter site, they may have a single skid that -- single measurement and separator skid combined, or they might have a separate separator skid and two separate meter run skids combined or together, as well as an instrumentation skid. And the	Topic 2: Meter Receipt Stations	

NATURAL LANGUAGE PROCESSING TOOLS

Topic Modelling

9758. **MR. SAWYER:** And as a result of that, NGTL did not undertake any cumulative effects assessment for the new meter stations, and instead relied on the cumulative effects assessment that was filed as evidence at the original NMML applications; is that also not correct?

--- (A short pause/Courte pause)

9759. **MR. MacMICHAEL:** I'd have to disagree with you there, Mr. Sawyer. It was not as a result of specifically that one determination. Through the effects assessment process, as guided by the NEB Filing Manual, we looked at all the valued components to consider what residual effects were possible as a component of the project effects after applying those mitigations.

9760. When it comes to cumulative effects, that was one component of how we determined whether or not to proceed further down the cumulative effects assessment process as laid out in Section 5.6 of the ESA.

9761. **MR. SAWYER:** So let's just be clear here, Mr. MacMichael, and let's just talk about caribou or wildlife for the moment. Your conclusion was, A, that there was no residual impact; correct?

9762. **MR. MacMICHAEL:** No, that they were negligible and that they were not significant.

9763. **MR. SAWYER:** Okay. And then you from that decision went to making a decision about cumulative effects that because they were negligible cum effects -- conducting a cum effects assessment was not necessary. Is that not also correct?

9764. **MR. MacMICHAEL:** No.

Topic 4: Cumulative effects assessment

Topic 3: Caribou population

Topic 4: Cumulative effects assessment

NATURAL LANGUAGE PROCESSING TOOLS

Topic Modelling and Sentiment Analysis

Topic 2: Meter Receipt Stations

8661.	**MS. HUME:** Thank you. But you do agree that workspace that was temporary in nature for the construction of the pipeline will now be permanently disturbed by the meter stations? It may be a small area, but there will be permanent disturbance that wasn't previously contemplated.	Concerns
8662.	**MR. PLATT:** Of the temporary workspace areas, there is a small footprint that will be permanent, but the nature of those disturbances are identical to those meter stations that were approved and are no longer being constructed as part of this project.	Neutral
8663.	**MS. HUME:** So you agree that the new meter stations will have permanent impacts that weren't contemplated as part of the initial pipeline application?	Concerns
8664.	**MR. PLATT:** The new meter stations are new applications for facilities connecting to the pipeline.	Neutral
8665.	**MS. HUME:** And they will permanently disturb new land?	Concerns
8666.	**MR. PLATT:** Yes, they will. But, again, the aggregate disturbance of the project -- the overall project is significantly reduced.	Neutral
8667.	**MS. HUME:** Thank you.	

PROPOSED SOLUTION

Metadata

- Year
- Type of hearing
- Speaker names
- Organization names
- Indigenous participants

Individual Transcript

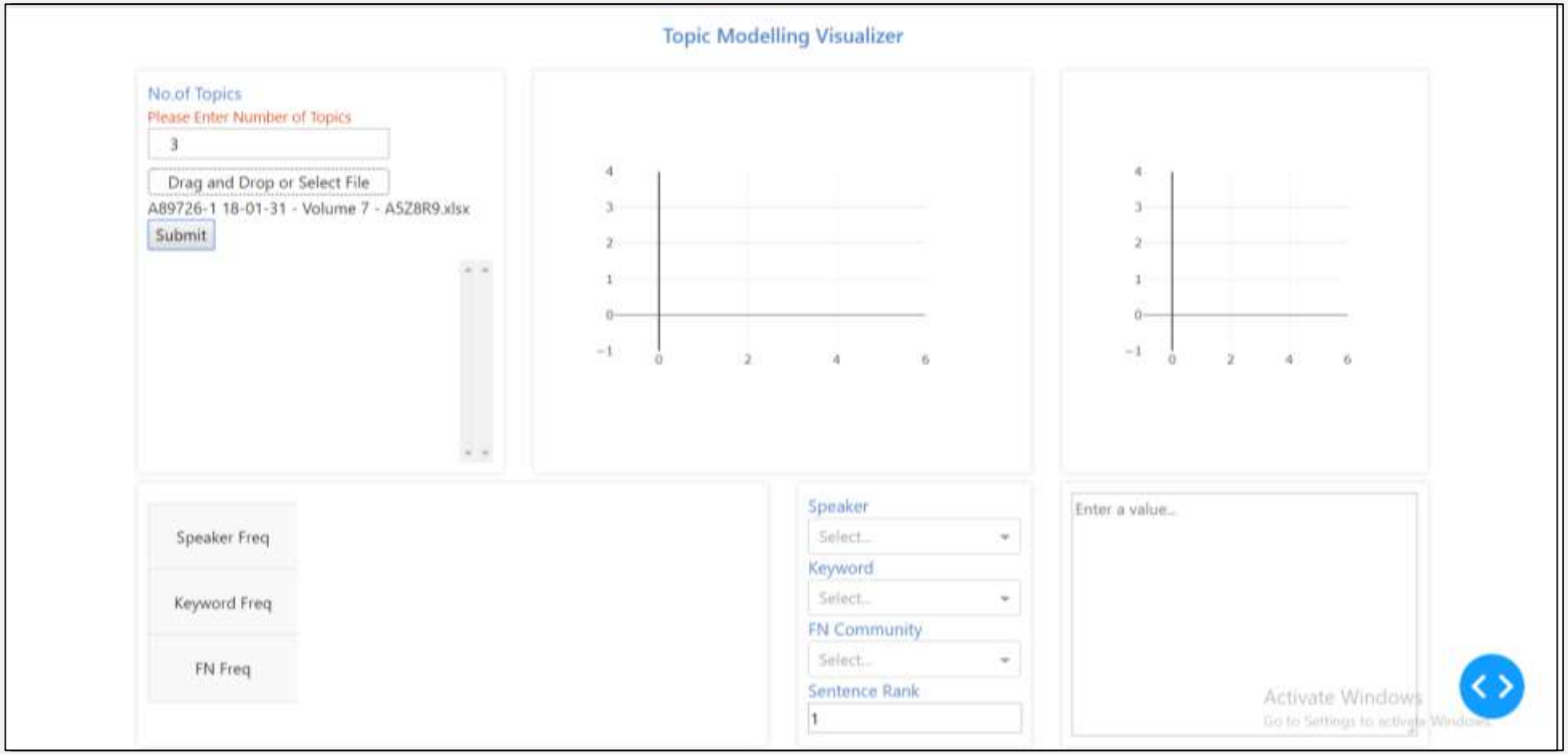
- LDA Topic Modelling
- Topic Visualization
- Top sentence
- Speaker frequency
- Sentiment analysis

Multiple Transcripts

- LDA Topic Modelling
- Topic Visualization
- Speaker/organization frequency across multiple transcripts
- Topic filtering by indigenous communities
- Sentiment analysis by topic and indigenous communities

INDIVIDUAL TRANSCRIPT VISUALIZER

Demonstration



Individual Transcript

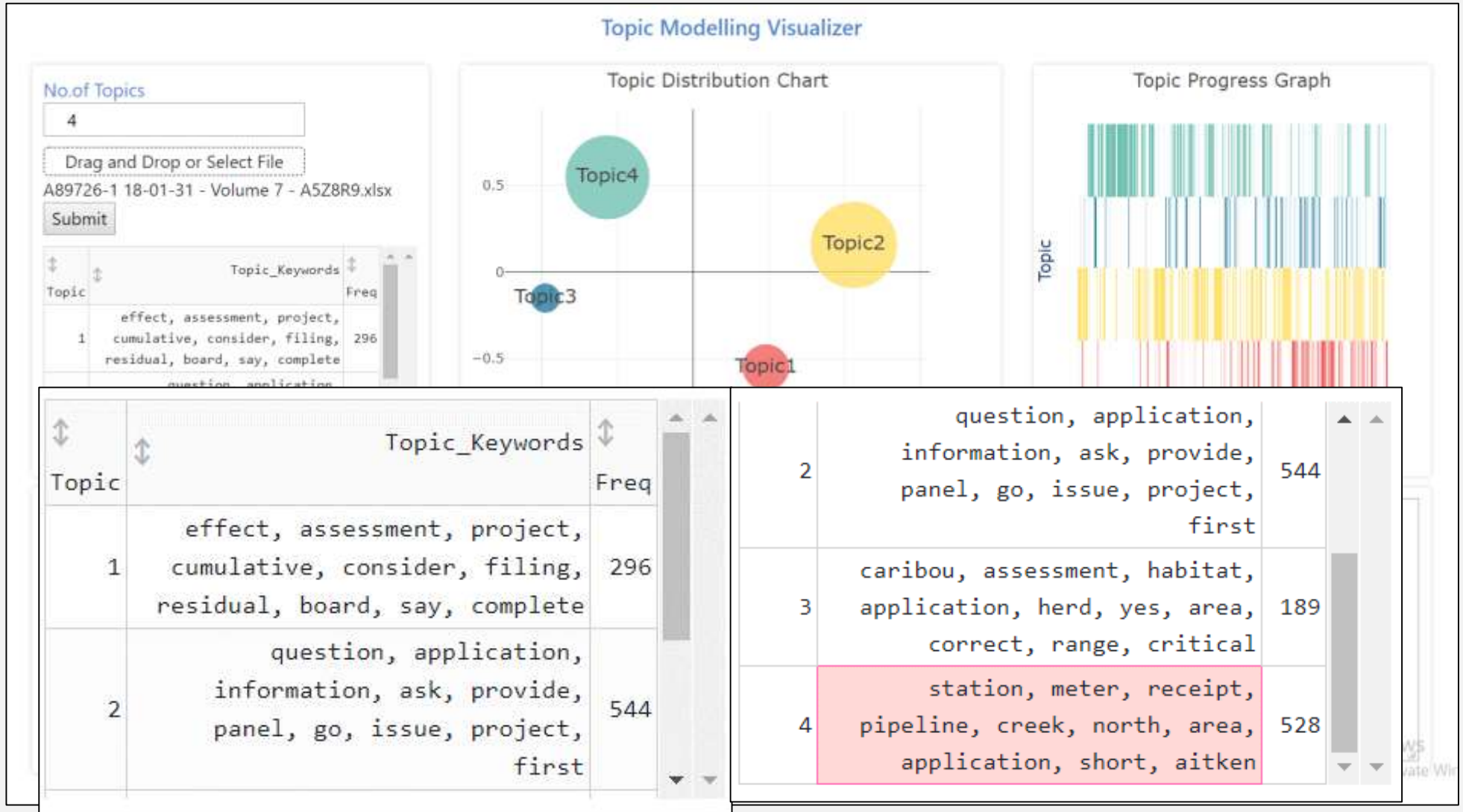
- LDA Topic Modelling
- Topic Visualization
- Top sentence
- Speaker frequency
- Sentiment analysis

INDIVIDUAL TRANSCRIPT VISUALIZER

Demonstration

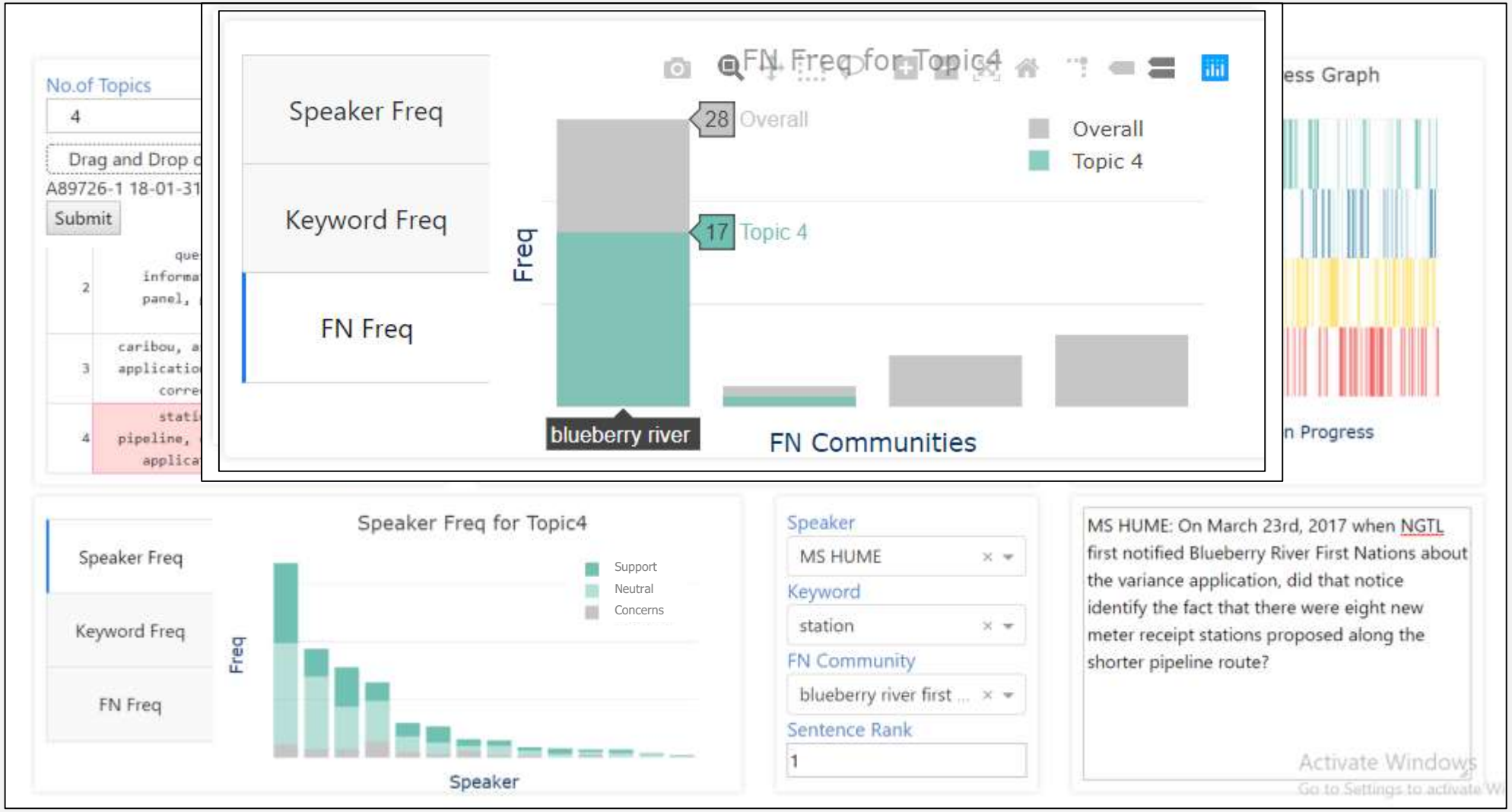
INDIVIDUAL TRANSCRIPT VISUALIZER

Demonstration



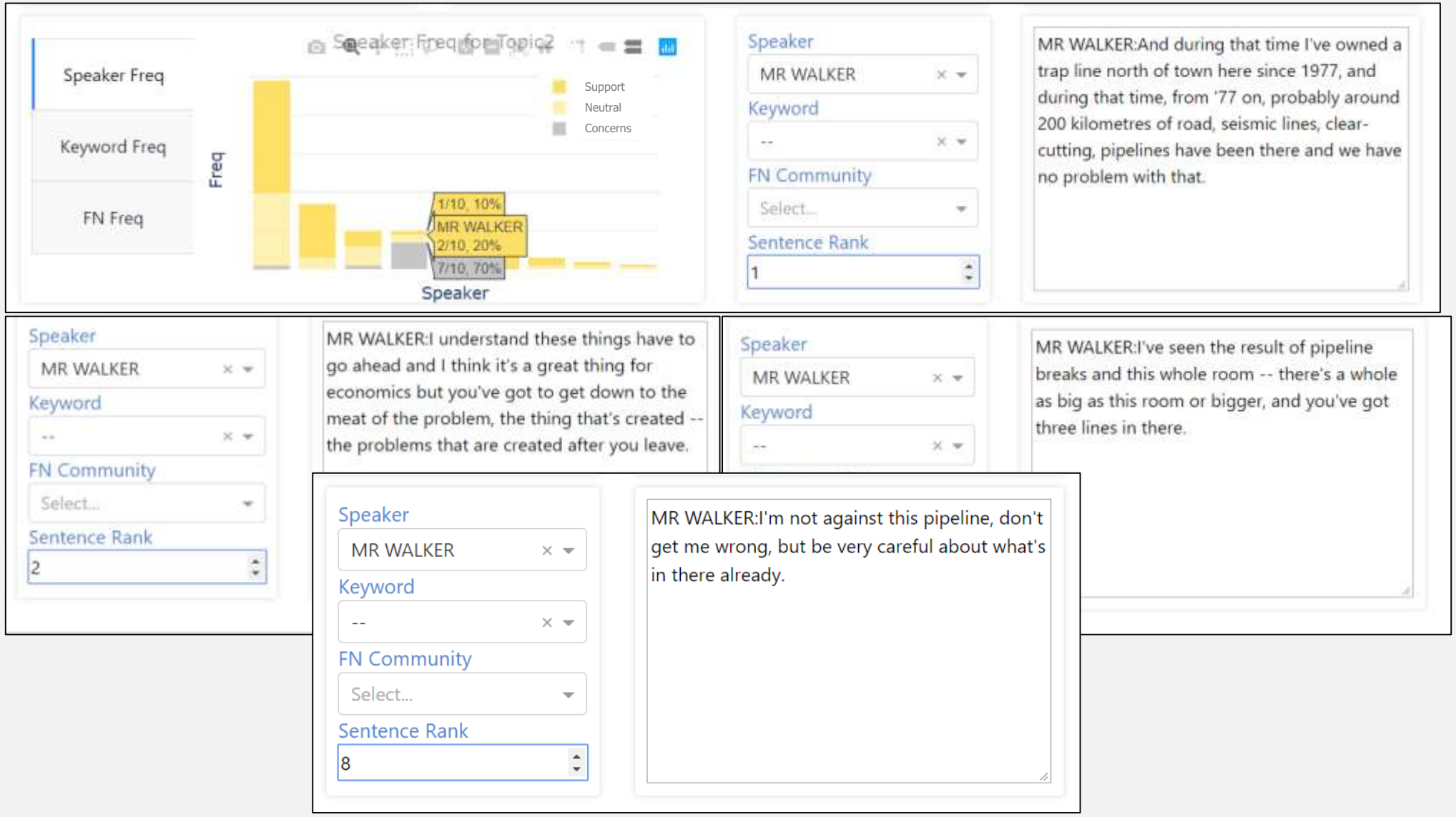
INDIVIDUAL TRANSCRIPT VISUALIZER

Demonstration



INDIVIDUAL TRANSCRIPT VISUALIZER

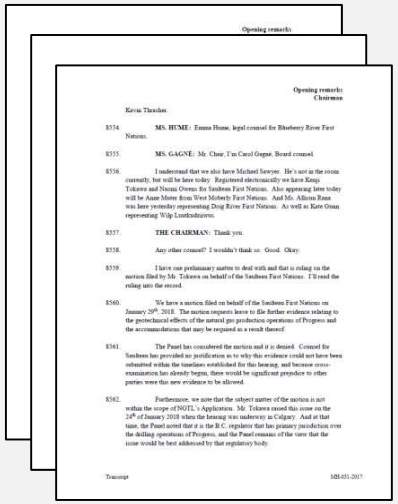
Demonstration



MULTI TRANSCRIPT VISUALIZER

Workflow; 2-step LDA Modelling

Step 1



Opening remarks

Opening remarks
Chairman

Kevin Thrasher.

8554. **MS. HUME:** Emma Hume, legal counsel for Blueberry River First Nations.

8555. **MS. GAGNÉ:** Mr. Chair, I'm Carol Gagné, Board counsel.

8556. I understand that we also have Michael Sawyer. He's not in the room currently, but will be here today. Registered electronically we have Kenji Tokawa and Naomi Owens for Saulteau First Nations. Also appearing later today will be Anne Mister from West Moberly First Nations. And Ms. Allison Rana was here yesterday representing Doig River First Nations. As well as Kate Gunn representing Wilp Lusstudzinsvu.

8557. **THE CHAIRMAN:** Thank you.

8558. Any other counsel? I wouldn't think so. Good. Okay.

8559. I have one preliminary matter to deal with and that is ruling on the motion filed by Mr. Tokawa on behalf of the Saulteau First Nations. I'll read the ruling into the record.

8560. We have a motion filed on behalf of the Saulteau First Nations on January 29th, 2018. The motion requests leave to file further evidence relating to the geotechnical effects of the natural gas production operations of Progress and the accommodations that may be required as a result thereof.

8561. The Panel has considered the motion and it is denied. Counsel for Saulteau has provided no justification as to why this evidence could not have been submitted within the timelines established for this hearing, and because cross-examination has already begun, these would be significant prejudice to other parties were this new evidence to be allowed.

8562. Furthermore, we note that the subject matter of the motion is not within the scope of NGTL's Application. Mr. Tokawa raised this issue on the 24th of January 2018 when the hearing was underway in Calgary. And at that time, the Panel noted that it is the B.C. regulator that has primary jurisdiction over the drilling operations of Progress, and the Panel remains of the view that the issue would be best addressed by that regulatory body.

Transcript MH-031-2017

All transcripts (1994-2018)

Train LDA Model to identify topics on a sentence-level

55 topics

7 categories (manually labelled based on keywords and sentences)

1. Pipeline
2. Information/check
3. Economic/financial
4. Land/property
5. Natural gas
6. Concerns
7. Administrative

Outcome: (sentence id, category id)

MULTI TRANSCRIPT VISUALIZER

Workflow; 2-step LDA Modelling

Step 2

Trained Categories

1. Pipeline
2. Information/check
3. Economic/financial
4. Land/property
5. Natural gas
6. Concerns
7. Administrative

→ Train LDA Model to identify topics on a sentence-level for each category (e.g. Pipeline)

→ Obtain detailed topics (manually labelled)

→ Further categorize sentences to new topics

Repeat for all 6 categories (excluding administrative)

MULTI TRANSCRIPT VISUALIZER

Categories and Topics

Pipeline [18 topics]:
1. risk scenario
2. price, credits, and rates
3. reserve areas and lands
4. oil spill and environment
5. water and life

…

Information/check [12]:
1. measurement check
2. local plans and specie risk analysis
3. pipeline safety management and water
4. concerns: land, community and traditional knowledge
5. board jurisdiction

…

Economic/financial [13]:
1. customer needs and requirements
2. firm service: contract and shipper
3. toll and cost
4. cost-of-service rate and standards
5. project application and board decision

…

Land/property [13]:
1. land, property, and caribou protection
2. field production and development
3. traditional use studies
4. winter scenario: production, price and ecosystem
5. spill, watercourse, and fisheries

…

Natural Gas [4]:
1. producer and contractor
2. plants and production capacity
3. gas flow, market, and costs
4. system design and utilization

Concerns [13]:
1. market and price determination
2. treaty rights and environment
3. financial risk and return
4. learning traditional oral knowledge
5. people and local community

…

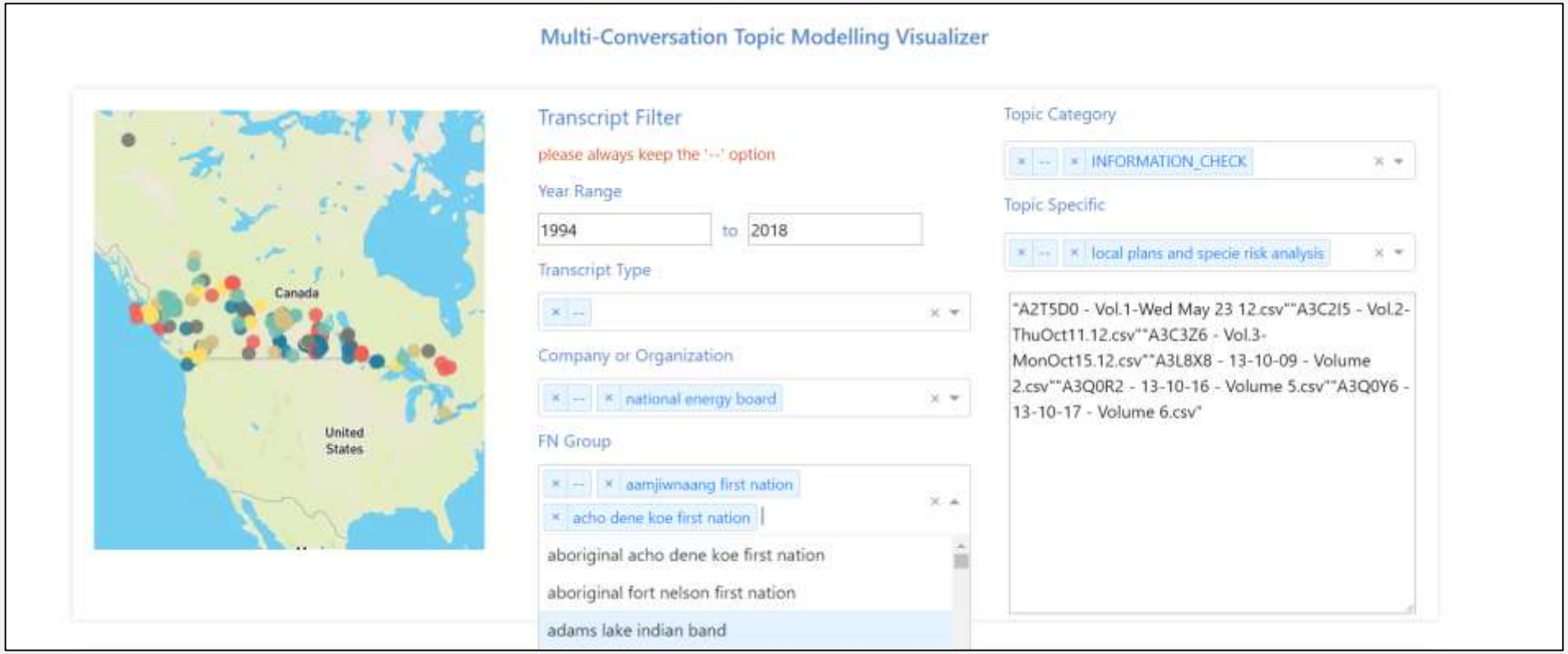
MULTI TRANSCRIPT VISUALIZER

General Applications

1. Assign each sentence to a dominant category and dominant topic
2. Get to know key themes for a single transcript
3. View temporal distributions of categories and topics on a timeline
4. Link themes to indigenous communities and other organizations
5. Compare thematic patterns in transcripts with political and institutional changes in reality

MULTI TRANSCRIPT VISUALIZER

Demonstration



Transcript Filter

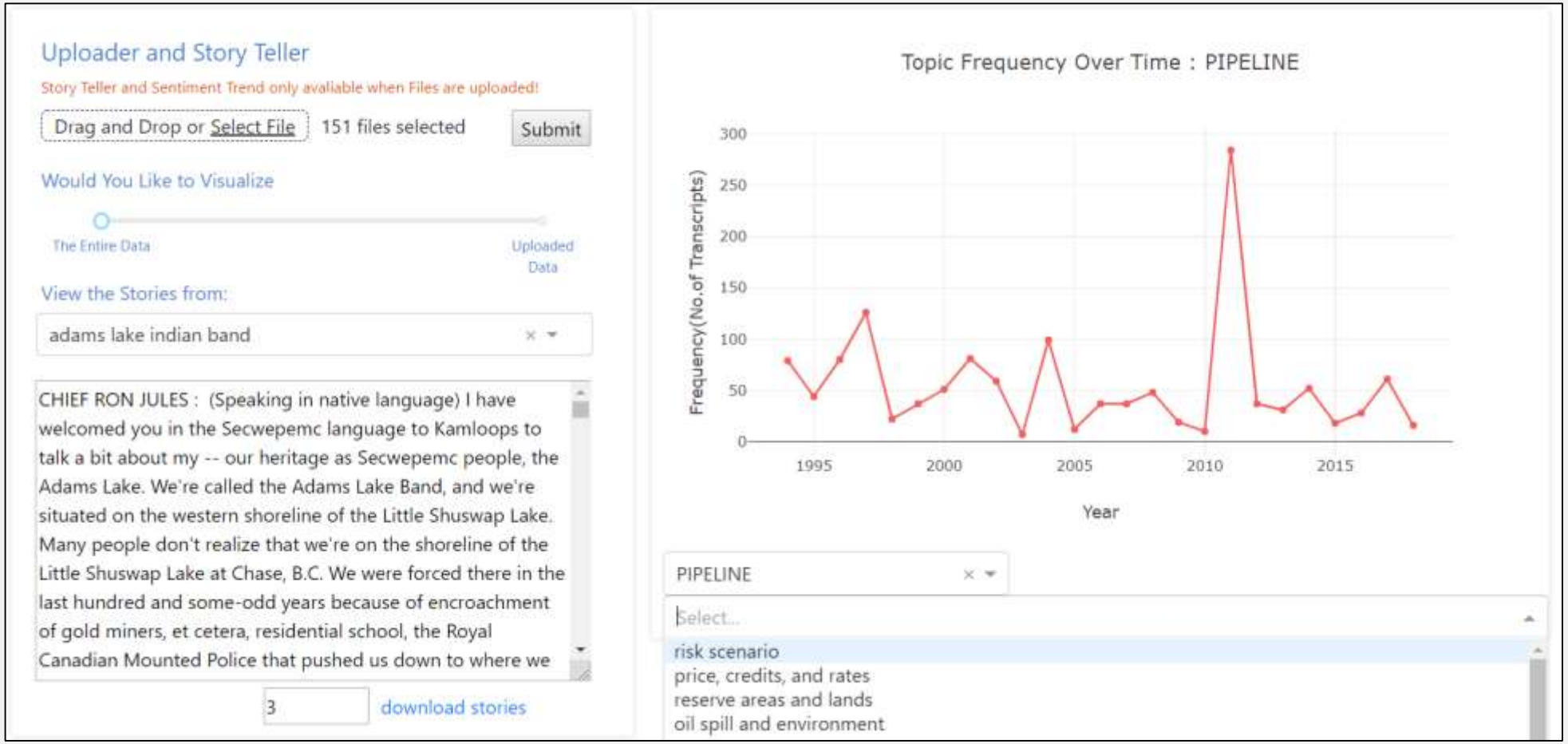
- Year
- Type of hearing
- Organization names
- Indigenous participants

Model on Entire Data

- General topics
- Get topics for each transcript

MULTI TRANSCRIPT VISUALIZER

Demonstration



Stories

- Upload Multiple File
- Get Indigenous Stories

Topic-based

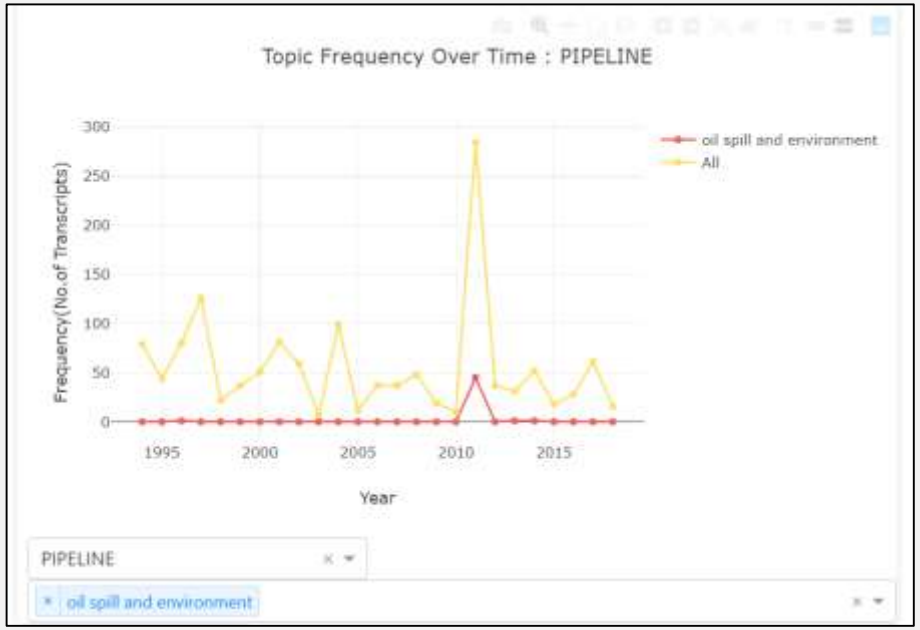
- Time-Series
- Sentiment

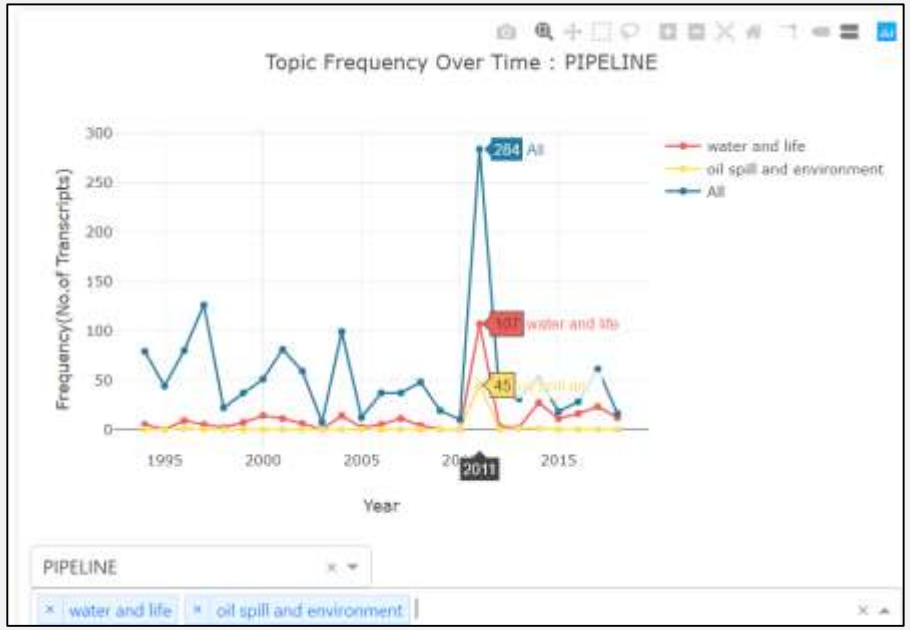
Community-based

- Topic distribution
- Time-Series
- Sentiment

MULTI TRANSCRIPT VISUALIZER

Demonstration





Stories

- Upload Multiple File
- Get Indigenous Stories

Topic-based

- Time-Series
- Sentiment

Community-based

- Topic distribution
- Time-Series
- Sentiment

2011 ALBERTA PIPELINE LEAK

BBC Sign in | News | Sport | Reel | Worklife | Travel | Future

NEWS

Home | Video | World | US & Canada | UK | Business | Tech | Science | Stories

US & Canada

Oil spill in Canada's Alberta 'biggest in 35 years'

4 May 2011 Share

A major oil pipeline leak in Alberta, Canada, is the largest the region has seen in 35 years, officials say.

Around 28,000 barrels of oil have spilled from a rupture in the Plains Midstream Canada Rainbow pipeline in northern Alberta.

CBC | MENU

news | Top Stories | Local | The National | Opinion | World | Canada

Edmonton

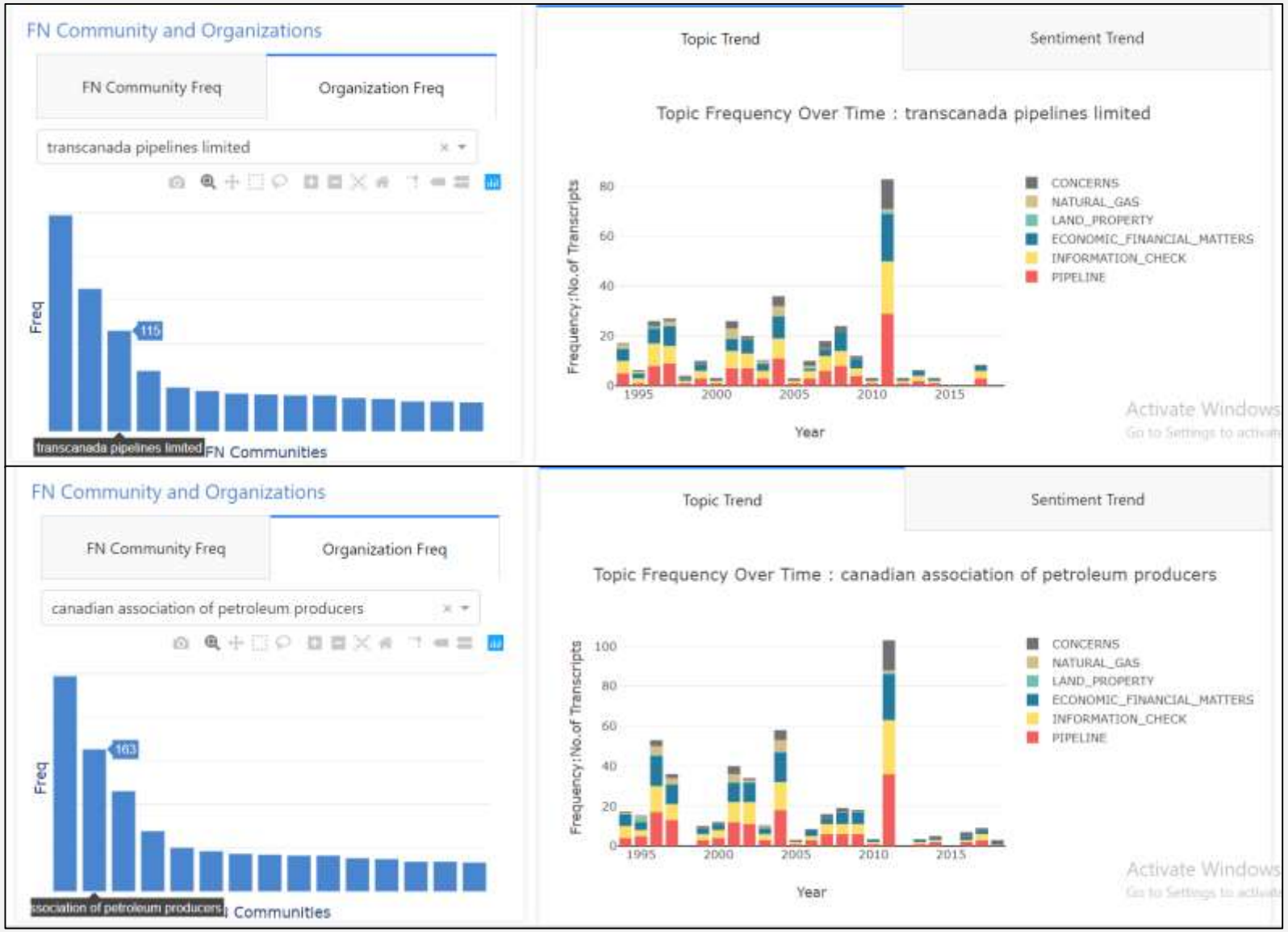
2nd largest pipeline spill in Alberta history leads to charges

Plains Midstream faces 3 charges in massive pipeline leak near Little Buffalo in 2011

CBC News · Posted: Apr 26, 2013 11:54 AM MT | Last Updated: April 26, 2013

MULTI TRANSCRIPT VISUALIZER

Demonstration



Stories

- Upload Multiple File
- Get Indigenous Stories

Topic-based

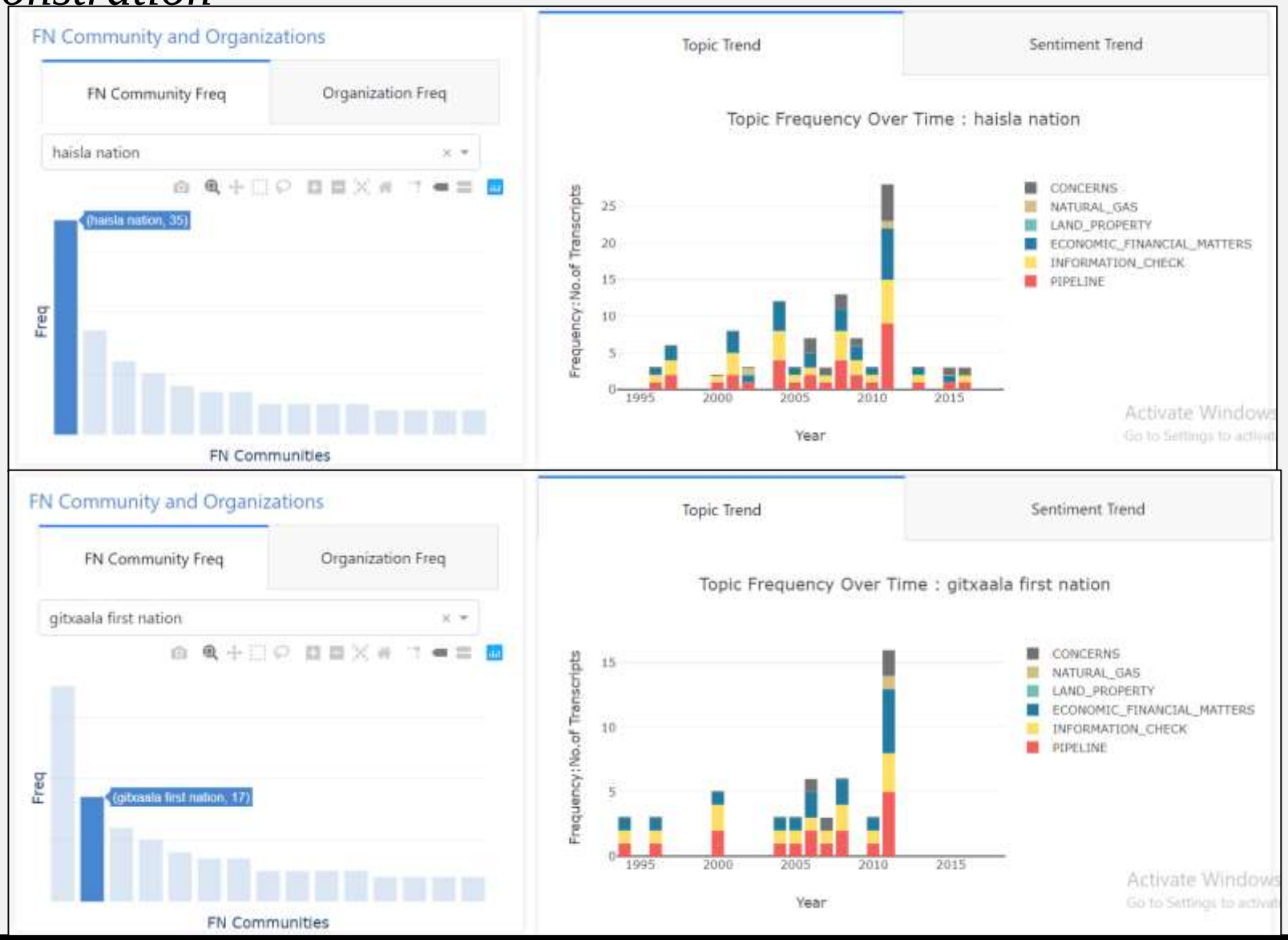
- Time-Series
- Sentiment

Community-based

- Topic distribution
- Time-Series
- Sentiment

MULTI TRANSCRIPT VISUALIZER

Demonstration



Stories

- Upload Multiple File
- Get Indigenous Stories

Topic-based

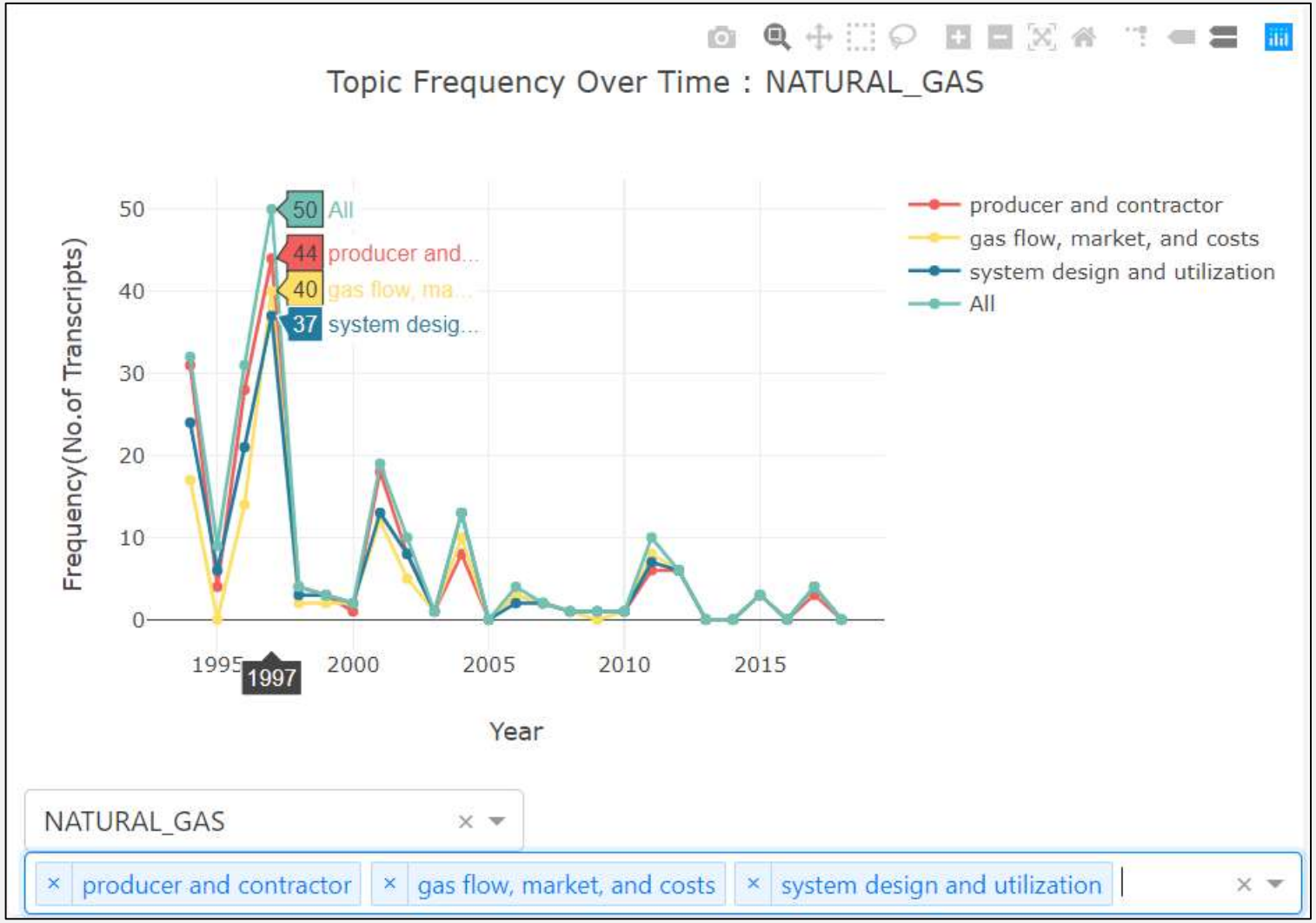
- Time-Series
- Sentiment

Community-based

- Topic distribution
- Time-Series
- Sentiment

MULTI TRANSCRIPT VISUALIZER

Demonstration



1997 INCREASED NATURAL GAS PRODUCTION

THE CANADIAN ENCYCLOPEDIA

EDUCATORS TIMELINES COLLECTIONS BROWSE SEARCH

CONTENT

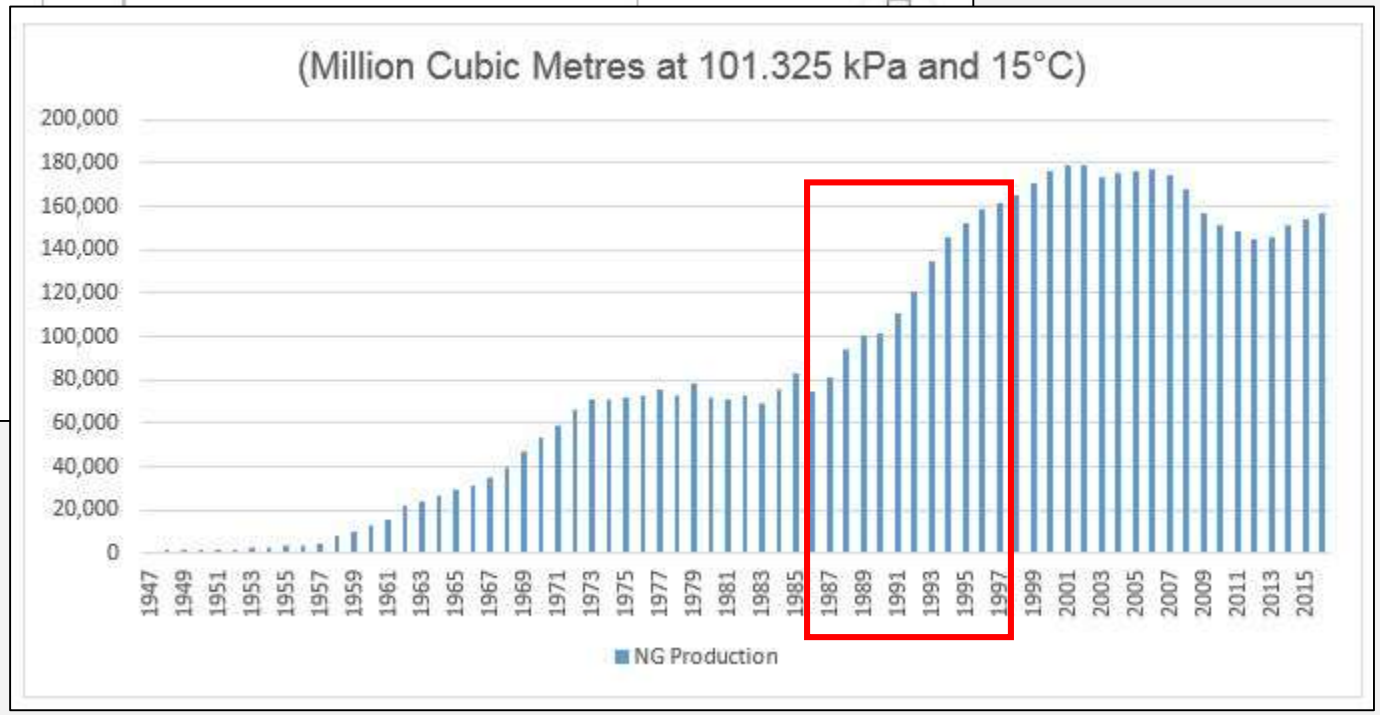
ARTICLE INDEX

- What Is Natural Gas?
- Where Is Natural Gas Found in Canada?
- Conventional and Unconventional Natural Gas
- History of Natural Gas in Canada
- Timeline: Canadian Natural Gas Industry
- How Is Natural Gas Produced?
- How Is Natural Gas Transported in Canada?
- Importing and Exporting Natural Gas
- How Is Natural Gas Stored?
- Projected Growth of Natural Gas
- Advantages and Challenges
- Glossary

RESOURCES

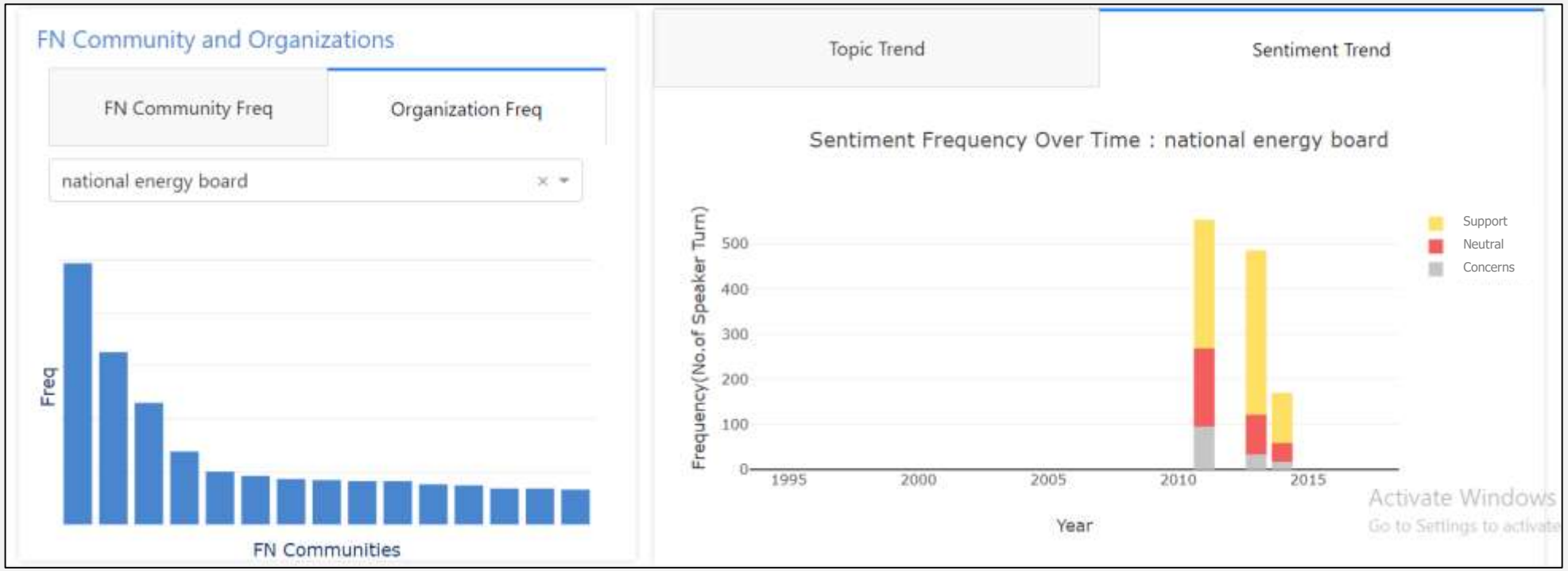
- Recommended

	The "Pre-build" pipeline originating in Caroline, Alberta, is completed, transporting natural gas from Western Canada.
1986	Discovery of the Caroline Swan Hills natural gas field.
1997	Production began at Newfoundland and Labrador's Hibernia offshore oil and gas field, discovered 18 years earlier in 1979.



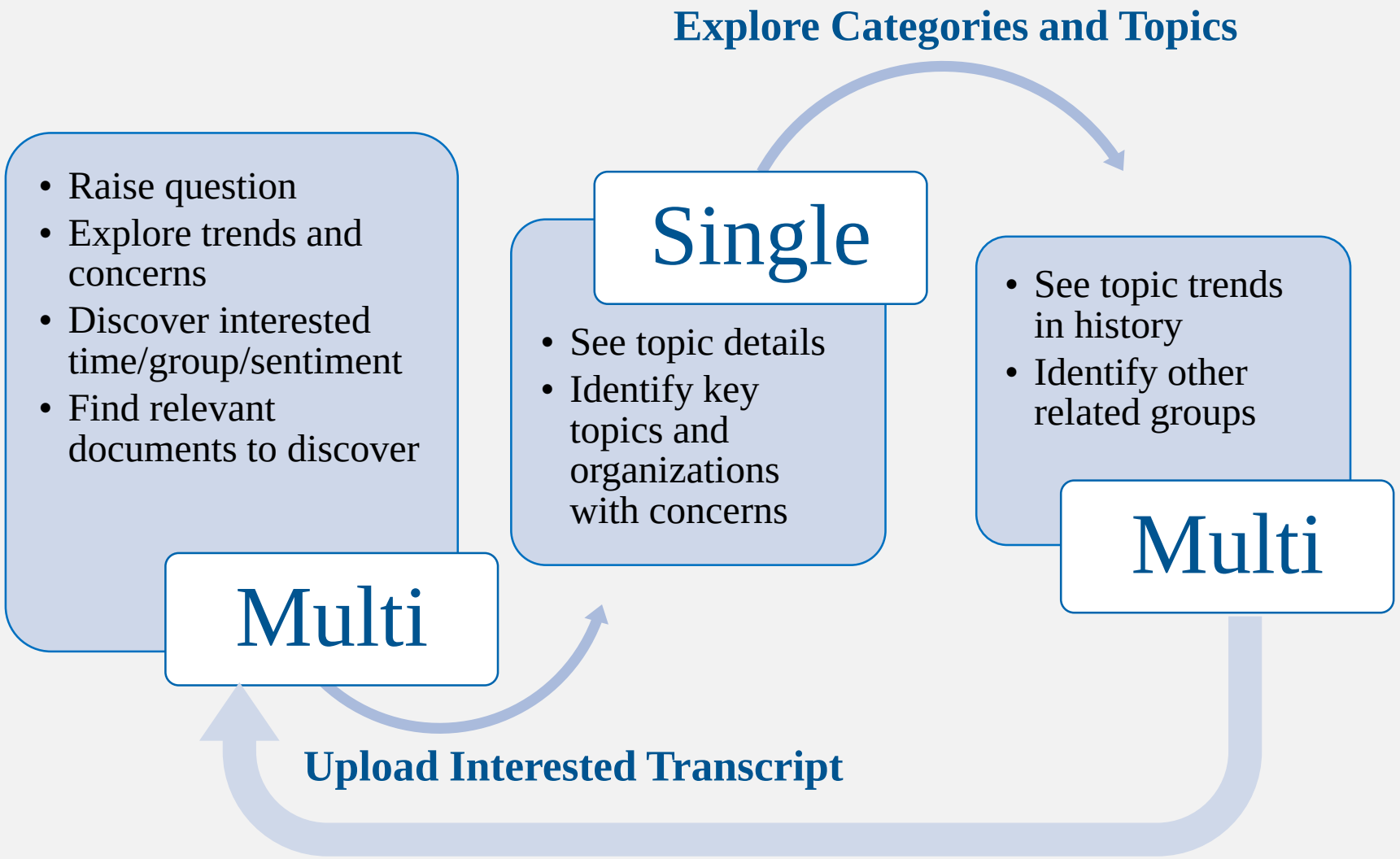
MULTI TRANSCRIPT VISUALIZER

Demonstration



SUMMARY

Multitranscript Visualizer to Individual Transcript Visualizer



CONCLUSION

Project Goals - Revisited

1. Over the last 60 years, who are the active participants in hearings?
2. What are the major concerns and interests in these hearings?
3. What are their attitudes towards these issues?
4. How can we preserve indigenous stories/testimonies from these hearings?

1. Now searchable through metadata
2. Able to view important topics and representative sentences/ statements through individual/ multi-transcript visualizer. Also able to view trends across time and communities/organizations.
3. Able to view sentiments of different communities/ organizations towards different topics
4. Visualizers can identify important sentences/statements by indigenous communities

CONCLUSION

Other Benefits and Future Work

Benefits

- Interrelated web applications that support one another (and are also subsequently supported by the clean metadata)
- Foundation and framework established for analysing large number of hearing transcripts

Future Work

- Continue updating database using provided metadata format
- Adopt other NLP methods to improve analysis (e.g. LSI, dialogue acts)
- Regression on hearing results using parameters created from topic modelling

THANK YOU

QUESTIONS?