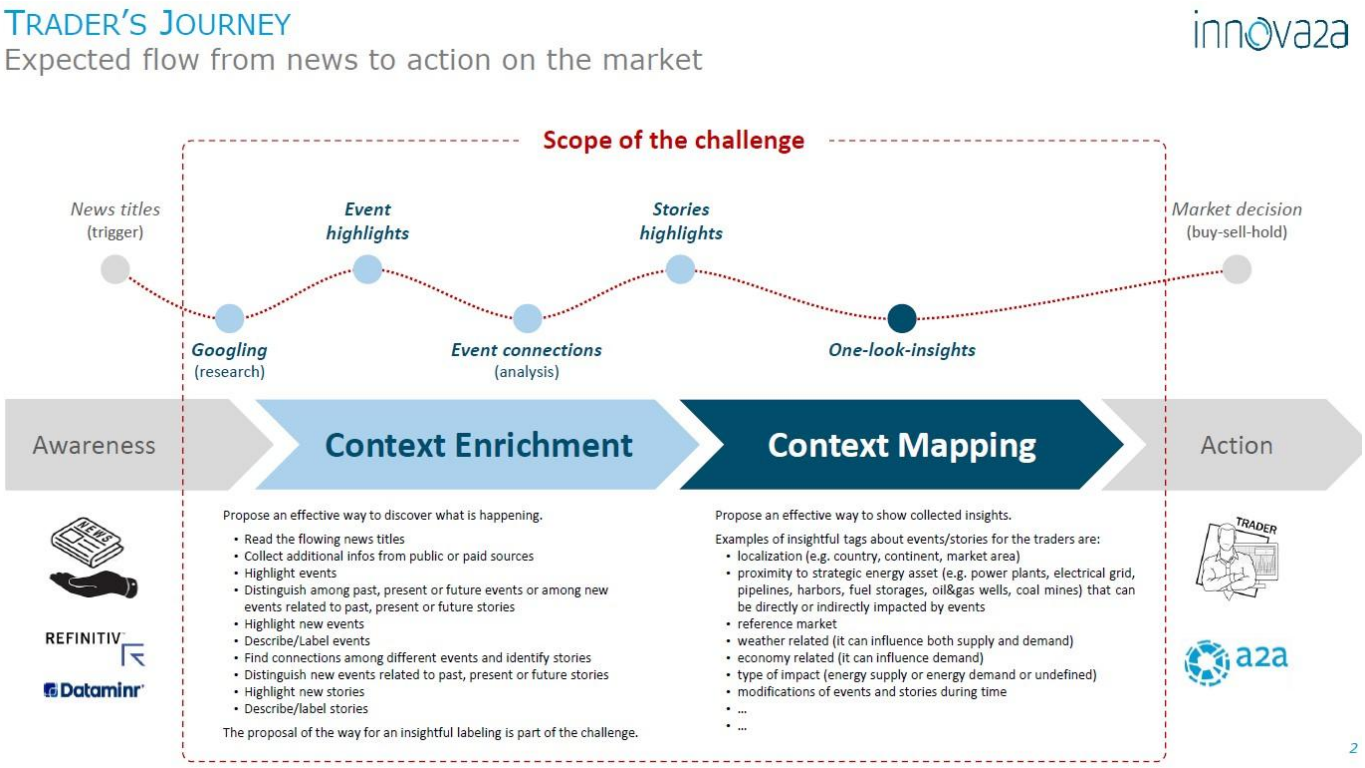
> Develop and implement an algorithm that ingests news titles, identifies relevant events, collects additional information about those events, and then effectively presents the collected insights using easy to understand visualizations.



Problem Overview

ABSTRACT

Staying informed about events that impact markets and business is critical to success in any field, and especially so for a Trader. With the ever-increasing flow of news and information over multiple channels and forms of media, keeping up-to-date and informed about the latest events is both easier and more difficult. Easier because news can be obtained anywhere at any time, but more difficult due to the need to sort and sift through the vast amount of incoming information in order to extract and contextualize valuable intelligence. The Key Stakeholders are interested in exploring new tools, techniques, or methods to reduce the effort required to process the news and aid in turning the flowing news into insightful intelligence about events that impact their business. Proposed solutions should focus on both **automated extraction and effective visualization of insightful events from news flows** and be presented in the form of an implemented algorithm.

BACKGROUND

News is an important source of information to improve trader understanding of what is happening on the markets and to fine tune trading strategy. Several news providers such as Refinitive and Dataminr already gather news and filter the massive flow of data supplying trading desks with the news items based on selected topics of interest. However, **traders are still required to read, verify and analyze the subset of news to highlight the meaningful events**. In fact, before taking actions on the markets, a trader must:

- Read the flowing news to highlight single events.
- Highlight a new event.
- Group the news related to the same event.
- If necessary, collect additional information to check validity.
- Contextualize events by correlating events belonging to the same story.
- Map events and stories and categorize by type:
- Localization – country, continent, market area, etc.
- Define proximity to energy assets such as power plants, electrical grid, pipelines, harbors, fuel storages, oil and gas wells, coal mines, etc. that can be directly or indirectly impacted by events.
- Reference market.
- Research weather related concerns (it can influence both supply and demand).
- Research economic related concerns(it can influence demand).
- Classify the impact (energy supply or energy demand or undefined).
- Understand evolution of events and stories through time.
- Make synthesis and take the best actions on the markets (buy/sell/stay)

Unfortunately, traders are often busy in trading desk operations, and have no time to continuously process the news. Therefore tools to automate the process outlined above would prove extremely useful in maximizing efficiency and outcome for the trading desk.

DATA

We are responsible for finding and utilizing data sources for the context enrichment and mapping required in this Challenge. The only input provided by the Stakeholder will be titles of news stories filtered by keywords of interest.

The Challenge

As illustrated in the scheme above, the Stakeholder is interested in exploring tools, techniques, or methods that can support their trading activities by turning the flowing

news into insightful events and accelerating context comprehension and decision-making. **The objective is to support decision making in real time helping traders to manage and classify the flow of events and rapidly understand and react to what is happening**. **The target is correct highlighting, validation, and classification of events and stories**. **Events are the single facts reported by the flow of titles provided by news providers** (such as Refinitive and Dataminr utilized by the Stakeholders), **while stories are sets of connected events.** For example, the story might be about gas supply interruption in central Europe and such a story includes multiple events ranging from an explosion that damaged a pipeline to the pipeline repair and its restoration to service. When classifying events it is important to know the event time stamp (different from the title time stamp) and if:

- The event is old it means it was already highlighted
- The event is new but belongs to an old story that was started by a previous event
- The event is new and does not belong to an old story it starts a new story but must be validated using validation criteria proposed by the Solver

We are asked to **develop and implement an algorithm that ingests news titles, identifies relevant events, collects additional information about those events, and then effectively presents the collected insights using easy to understand visualizations.** There are two main focus activities that submissions for the Challenge must address using data sources found by the Solver – context enrichment, and context mapping – as described below. We are responsible for finding and utilizing any data source used by the proposed algorithm.

Stages and Requirements

Context Enrichment, Context Mapping, Data Visualization

1. Context Enrichment

Processing

Propose and implement an effective way to **discover what is happening**:

- Read the flowing news titles and identify relevant news items.
- Collect additional information from public or paid sources.
- Highlight events.
- Distinguish among past, present, or future events or among new events related to past, present or future stories.
- Highlight new events.

- Describe/Label events.
- Find connections among different events and identify stories.
- Distinguish new events related to past, present or future stories.
- Highlight new stories.
- Describe/label stories. (A proposal for a way of insightful labeling is part of the discussion).

Output

- Title
- Title timestamp
- Event name
- Event timestamp
- Event status (validated/not validated)
- Event type (new/old)
- Story name
- Progressive story timestamp
- Story status (validated/not validated)
- Story type (new/old)
- Event/Story link

2. Context Mapping

Processing

Propose an effective way to **visualize and show collected insights**. Information collected during *context enrichment* may also be used for context mapping and proposed algorithms must be able to extract tags and features that fully describe an event or story and its evolution over time. **Commonly used tags** to judge the impact of an event in the energy sector are:

- Localization (for example, country, continent, market area).
- Proximity to energy assets (for example, power-plants, electrical grid, pipelines, harbors, fuel storage, oil and gas wells, coal mines) that can be directly or indirectly impacted by events.
- Reference market.
- Weather related (it can influence both supply and demand).
- Economy related (it can influence demand).
- Type of impact (energy supply or energy demand or undefined).
- Modifications of events and stories during time.

Output

- Event name
- Event brief description
- List of tag types considered to describe an event
- List of tags that fully describe the event and enable analysis
- Story name
- Story brief description at each timestamp
- List of tag types considered to describe a story at each timestamp
- List of tags that fully describe the story and enable analysis at each timestamp

Data sources to enable tagging of events or stories must be proposed by the Solver. Using the previous example of a gas pipeline explosion the identification of pipelines and markets impacted by the explosion is within the scope of the algorithm. The tags listed above are necessary but may not be enough to provide a comprehensive description of events and stories and proposed algorithms should have the **capability to identify and optimize tags to fully describe an event or story**. Algorithms should be able to **create and update a database of energy assets identified in the enrichment phase for use in the tagging and mapping phase**. Examples of public sources of data about electricity and gas infrastructure are https://transparency.entsog.eu/ and https://transparency.entsoe.eu/.

3. Data Visualization

Output

- Interface that provide a synthesis of events/stories
- Interface that enable monitoring of events/stories during time
- Description of interface features to configure and adapt data visualization
- Concept/Mockup can be accepted

Hypothetical Sudo Code

Processing

```
get title
get article source
extract meta from article
if meta matches existing event meta stop here
create new event
        set meta
        set event name
        set event timestamp
```

```
        validate event
        set event type
        set event info
        set event context
connect event to story
        (match event data to event data to determine events in story)
        if event matches existent story
                add event to story
                if event is update to story
                        create update in story
        else create new story
```

Representation (in real time; newest first)

```
display stories
        display story data
display events
        display event data
        display parent stories
```

Data model

```
event {
        meta : meta
        event name
        event timestamp : past / present / future
        event validated : boolean
        event type : new/old
        event info : info
        context : context
}
meta {
        title
        timestamp : past / present / future
        url
}
story{
        meta : meta
        updates : [update]
        story validated : boolean
        story type : new/old
```

```
        events : [event]
        context : context
}
story update{
        timestamp
        info : info
}
info {
        name
        description / label
        summary
        tags : []
}
tag {
        name
}
context {
        localization : location
        impacted assets : [impacted assets]
        reference market : related market from list of markets
        relations : [relation]
}
location {
        geotag
        country
        continent
        local market
}
proximity {
        asset
        impact
}
asset {
        name
        type : powerplant,electrical grid,pipelines,harbors,fuel storage, oil and
gas wells,coal mines
        location : location
}
relation {
        name : weather / economy
```

```
        impact : [supply/demand]
}

--------------------------------------------------------------------------------
### theoretical

To validate new event keep a database of trusted sources. Allow trader to
"validate" or "invalidate" an event as proof to their trust for said source.

source {
        name
        validations : number
        invalidations : number
}
```

Evaluation Procedure and Criteria

The solution must be an algorithm that meets the following technical requirements:

> Automatically accomplish **Context Enrichment**, **Context Mapping**, and **Data Visualization** as described above for article titles from a news feed. The algorithm must replicate and automate the process outlined in the above scheme and visualize the map of insights.

Metrics used to evaluate solutions will be based on the following criteria:

- Number of highlighted events based on the provided dataset titles
- Number of titles per event.
- Number of events per story (or, alternatively, number of links among highlighted events).
- Time necessary for context enrichments (update of “context enrichment” file/database).
- Time necessary for context mapping (update of “context enrichment” file/database).
- Number of data sources alternative to search engines (for example google) used for context enrichment/mapping.
- Percentage of data sources that are open source or feely available (among the proposed alternative sources).
- Unsupervised-learning capability to optimize event/story tagging/featuring (Yes/No).
- Distribution of tags in term number and type per events/stories.
- Usability of data tagging strategy for classification/clustering (qualitative).

- Effectiveness of data visualization strategy (qualitative).
- Flexibility in data visualization configuration (qualitative).