What does the application do?

- Continuously collect multilingual media reports from web portals.
- Every 10 minutes, cluster the latest articles about the same event or subject (‘live clusters’).
- Identify and extract information on events of the types: violence, natural disaster and humanitarian crisis.
- Display the latest events on a map.
- Give access to extracted information and to full articles.

The processing chain

Extracted information

- **Victims**: Number, Status (injured, killed, kidnapped, ...), Description (e.g. 3 Palestinian women)
- **Perpetrators** (e.g. militant groups)
- **Place** of the event
- **Event type** and weapons (e.g. bomb, missile, knife)
- **Event types**:

<table>
<thead>
<tr>
<th>Pattern engine / Formalism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ExPRESS</strong> is a blend of JAPE (GATE) and XTDL (SProUT).</td>
</tr>
<tr>
<td><strong>LHS</strong> of the rule is a regular expression over flat feature structures.</td>
</tr>
<tr>
<td><strong>RHS</strong> specifies the output structure.</td>
</tr>
<tr>
<td>Allows variables, labels, functional operators, grammar cascading, ...</td>
</tr>
<tr>
<td>Multiple and nested labels (multiple actions)</td>
</tr>
<tr>
<td>Rule sample:</td>
</tr>
</tbody>
</table>

```
iquery-event -> (person-group $ [NAME #name1, NUMBER #num1]; injured1
|token & [SURFACE "and"]
| [person-group $ [NAME #name2, NUMBER #num2]; injured2
| injured-phrase $ [FORM "passive"]
|) event

-> injured1 victim $ [NAME #name1, NUMBER #num1];
injured1 $ [NAME #name1, NUMBER #num2];
injured2 $ [NAME #name1, NUMBER #num2];
event $ [person-group $ [NAME #name, NUMBER #num];
|name = Concatenate(name1,"and",name2)
|) event
```

Features

- **Shallow analysis**, simple (1 or 2-slot) extraction rules.
- **Information aggregation** at cluster level, resolving contradictory information (e.g. on number of victims).
- **Geo-tagging** of events: place name disambiguation using occurrence frequency in cluster, size of place and place name hierarchy (country, region, province, city).
- **Event classification** is based on keywords, an event taxonomy, Naïve Bayes classifiers and domain-specific rules.
- **Pattern learning**: seed patterns extract information (e.g. victims); algorithm searches for paraphrases within cluster, exploiting redundancy in news reporting.
- **Efficiency**: capable of processing large amounts of data in real time.