Automated journalism: an overview

Laurence Dierickx
Université Libre de Bruxelles
2018
What is automated journalism?

- A technological process: from data to text
- A subfield of computational linguistics (first experiences of NLG in 1970s)
- An editorial process: human intentions designed into computer processing

A matter of choices made by human beings
What is not automated journalism?

- A technology that will act by itself
- An uncontrolled process full of bias *(bias or errors are from humans or from bad data quality, a computer software never repeat the same mistake... if it is detected!)*
- A 3D robot even if the metaphor is widely used *(except in Asia)*
How does it work?

WHAT TO SAY?
HOW TO SAY IT?

Automated journalism: an overview
Laurence Dierickx
What are the requirements?

- Structured data
- Rights to use the data
- Reliable source
- Data must be up to date
- Good data quality = accurate, complete,…
- Relevance for a journalistic use

Automated journalism: an overview
Laurence Dierickx
What are the covered domains?

Limitations are consequences of the requirements. Everything cannot be automated!

Automated journalism: an overview
Laurence Dierickx
<table>
<thead>
<tr>
<th>Year</th>
<th>Source</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Google News</td>
<td>Outdated financial data published with the date of the day while it was much older, announcing the bankrupt of a company. Consequences: traders sell... (impact on markets).</td>
</tr>
<tr>
<td>2017</td>
<td>L.A. Times</td>
<td>Quakebot (2014): massive earthquake (magnitude 6.7), panic on social networks! This earthquake had happened well... but in 1925! Error came from wrong encoding (by a human) in the database (U.S. Geological Survey..)</td>
</tr>
<tr>
<td>2015</td>
<td>Associated Press</td>
<td>Netflix released second-quarter earnings at the same time as its stock underwent a 7-to-1 split. Value were no split: misleading interpretation (71, fall of individual share).</td>
</tr>
</tbody>
</table>
Which European media do use it?

PRESS AGENCIES
NEWSPAPERS
MAGAZINES
PURE PLAYERS

WE DON’T KNOW EVERYTHING!
CONFIDENTIAL DISCLOSURES
(Germany)

Readers are not ready?
Newsrooms are not ready?

Automated journalism: an overview
Laurence Dierickx
Which European media do use it?

Automated journalism: an overview
Laurence Dierickx
<table>
<thead>
<tr>
<th>Why do they use it?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>To speed up content production</strong></td>
</tr>
<tr>
<td>High speed level production: in France, more than 1 billion articles produced in one night (elections)</td>
</tr>
<tr>
<td><strong>To extend media coverage</strong></td>
</tr>
<tr>
<td>To cover events not covered before (f.e NTB sports coverage in Norway)</td>
</tr>
<tr>
<td><strong>To provide a service to the readers</strong></td>
</tr>
<tr>
<td>Real time information about hot topics</td>
</tr>
<tr>
<td><strong>To make something that was not done before</strong></td>
</tr>
<tr>
<td>Dealing with large amounts of data can be tricky. Automation can help in a fast and accurate way.</td>
</tr>
<tr>
<td><strong>To assist journalists</strong></td>
</tr>
<tr>
<td>By freeing them of repetitive and time-consuming tasks.</td>
</tr>
<tr>
<td><strong>Not to replace human journalists</strong></td>
</tr>
<tr>
<td>Nothing in this way is observed (employees) but there is no zero risk! Freelancers are more fragile.</td>
</tr>
</tbody>
</table>

**Automation systems can be easily multilingual**

Automated journalism: an overview
Laurence Dierickx
What are the obstacles for development?

1. Data
   Available, structured, reliable

2. Costs
   Between 20 & 50k *Le Monde* // projects supported by Google DNI (UK 708k, Sweden 46k) // public funds (Netherland, T.U, 700k)

3. Resources
   Need internal skills

Automated journalism: an overview
Laurence Dierickx
Who are the humans behind?

1. **IT service**
   Inside the media (f.e. NTB, Norway)

2. **Tech start-up (not claiming themselves as media cies)**
   Mostly in France (Syllabs) and in Germany (Retresco, AX Semantics)

3. **Journo-dev**
   Marginal cases
How do journalists react?

1. **Enthusiastic**
   Free-up time, repetitive and boring tasks for the machine

2. **Pessimistic**
   Fear that automated systems will replace journalists

3. **Fatalist**
   If automation can destroy jobs, it can also create new ones
Do the audiences make the difference?

According to studies, the answer is no but…

- audiences find generated contents less pleasant to read
- automated stories are considered as less well written
- in the same time, those stories are judged more accurate and reliable
Tool for journalists?

Case study (Belgium): Bxl’air bot

Automated journalism: an overview
Laurence Dierickx
Tool for journalists?

www.bxlairaibot.be (Alter Echos)

- Retrieve and store data about air quality in Brussels
- Real time open data (HTML) collected from different sources
- Provide first data analysis: text, charts, maps
- Monitoring (1 year), raw material for investigative work (causes & consequences)

Automated journalism: an overview
Laurence Dierickx
Tool for journalists?

Assets for the journalists

- Work by its own to make a long time job (gain of time)
- Accuracy (supposed a human control), objectivate by numbers (exceeding of EU norms)
- No need to get particular skills (at the opposite of datajournalism)
- Used by journalists only if they can make sense of it

Automated journalism: an overview

Laurence Dierickx
Thank you!

@ohmyshambles
www.ohmybox.info