Content management techniques and tools for fact-checking

Francois Goasdoue\textsuperscript{a,b}, Ioana Manolescu\textsuperscript{b,c,d}, Xavier Tannier\textsuperscript{d,e}

\textsuperscript{a} U. Rennes 1 \textsuperscript{b} Inria Saclay \textsuperscript{c} LIX (CNRS and Ecole Polytechnique)
\textsuperscript{d} U. Paris Saclay \textsuperscript{e} LIMSI (CNRS and U. Paris Sud)
Plan

1. Data journalism and fact-checking
2. ... are content management problems
3. Some works I have been involved to in this area
Data journalism

Investigative journalism based on complex and/or large data

http://abonnes.lemonde.fr/les-decodeurs/portfolio/2017/04/18/les-fractures-francaises-1-5-le-logement-les-raisons-de-la-crise_5112859_4355770.html
Data journalism

Investigative journalism based on complex and/or large data

http://abonnes.lemonde.fr/les-decodeurs/portfolio/2017/04/18/les-fractures-francaises-1-5-le-logement-les-raisons-de-la-crise_5112859_4355770.html
Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)
Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)
Fact-checking (since 1930 approx.)

**Fact-checking**: verification of facts mentioned in media content
- To protect media reputation and avoid legal action
- Verification supposes the existence of a reference dataset

“The day I became a fact-checker at The New Yorker, I received one set of red pencils […] for underlining passages on page proofs of articles that might contain checkable facts. […] confirmed with the help of reference books from the magazine’s library, including Merriam-Webster’s Geographical Dictionary, the New Grove Dictionary of Music and Musicians and Burke’s Peerage and Gentry.”

Fact-checking (2012 – ongoing)

Did Democrats Once Support Border Wall?

By Robert Farley  Posted on April 26, 2017

Mick Mulvaney made an apples-to-oranges comparison when he said he couldn’t understand why Democrats opposed supplemental funding for a border wall since many of them were for it back in 2006. Mulvaney is referring to the Secure Fence Act of 2006, which called for construction of 700 miles of fencing and enhanced surveillance technology, such as unmanned drones, ground-based sensors, satellites, radar coverage, and cameras. See Chuck Schumer and then-Sen. Barack Obama and Hillary Clinton were among a bipartisan majority that voted in favor of the legislation, and it was signed into law by President George W. Bush.

In a very general sense, the Democrats named by Mulvaney supported a bill to build more border fences.

Q: Did the Supreme Court rule that public schools cannot teach students about Islam? A: No. That false claim was spread by a network of fake news websites.

Not everyone agrees, however, that Democrats are not flip-flopping on the issue. Mark Krikorian, executive director of the Center for Immigration Studies, a think tank that advocates for lower immigration, said that because the public doesn’t know exactly what border barriers the Trump administration wants to build, Mulvaney’s statement is not an “exact” comparison. But, he said, to dismiss it simply on that basis would be “tendentiously literal.”

“The fact is that, other than the ‘Mexico will pay for it’ stuff, Trump is simply channeling the 2006 Secure Fence Act, and Schumer et al. who voted for it out of political calculation are indeed hypocrites for opposing the attempt to finally bring that law to fruition,” Krikorian told us via email.

At the surface level, it is true in a broad sense that Democrats including Schumer, Obama and Clinton have in the past supported border fencing. All three voted for the Secure Fence Act of 2006, and all three supported the 2013 Senate immigration overhaul that passed the Senate, and which called for tougher border security including some additional fencing. But to claim that those measures are the same as what Trump is proposing is a stretch.

Share The Facts

Mick Mulvaney
Director, Office of Management and Budget

“We don’t understand why the Democrats are so wholeheartedly against [President Trump’s border wall]. They voted for it in 2006.”

Fox News Sunday – Sunday, April 23, 2017

SHARE 
READ MORE

GOASDOUÉ, MANOLESCU, TANNIER / PARIS DB DAY
Washington Post’s TruthTeller (2013)

Discontinued

Video → Automated transcript → manual matching into FactCheck.org
Fact-checking is a content management problem.

**Media content**
- Claim to be checked (text or data)
- Human actors (journalists, experts, crowd workers)

**Media context**
- Verification tool (query, match, source search...)
- Reference information source 1
- Reference information source 2
- Reference information source n

Analysis result:
« True / rather true / rather false / false
See sources: [http://dataref.com... »
Fact-checking is a content management problem

- **Media content**
  - Claim extraction
  - Claim to be checked (text or data)
    - Human actors (journalists, experts, *crowd workers*)
    - Reconciliation, reputation

- **Media context**
  - Social network analysis
    - Verification tool (query, match, source search...)
    - Source selection
      - Reference information source 1
      - Reference information source 2
      - ... Reference information source n
    - Reference source construction, refinement, integration

Analysis result
« True / rather true / rather false / false
Fact-checking is a content management problem
A point of caution: it’s not just checking
Most aspects of modern reality are complex
From a journalistic perspective, explaining may be as important and useful as checking
Also: future is hard(er) to check
Toward automated fact-checking


Browser plug-in

Bringing up rich context for a Web page, from document and knowledge bases

**Source search**

Supported by template queries over XML documents and RDF graphs

« Second screen »
Content management tools for data journalism


So many data sources, so little time!

~ data lake

« Can we group tweets of politicians by political current and analyze their most frequent topics? »

« Can we classify articles by their main concept and do a tag cloud? »

Can we industrialize answering such requests?
Content management for data journalism


So many data sources, so little time!

Data journalism; reference source enrichment

Dataviz modules

Tatooine query engine

Custom processing
Classification, ...

Data sources

JSON sources (SOLR, Postgres...)
Full-text search Indexing

Relational sources (RDBMS)

SQL Indexing

RDF sources (Jena, Corese...)

SPARQL Reasoning

Streams (Twitter, RSS)

Subscribe Query

Demo
Modeling facts, statements and lies

Ongoing work with Ludivine Duroyon (U. Rennes 1) and Francois Goasdoué

We need to represent information such as:

1. Penelope worked for the National Assembly (NA) from 2002 to 2012
2. Her first work contract for the NA ran from 2002 to 2005 = in 2002 the employee database showed that she was going to work there from 2002 to 2005
   On Jan 21, 2017, he corrects himself to state that Penelope worked for the NA from 2002 to 2008.
4. On Jan 22, 2017, Le Canard Enchaîné wrote that François had stated on Jan 21 that Penelope worked for the NA from 2002 to 2008.
5. Charles and Marie both worked for the NA at some point, but Charles' contract was after Marie's (they never overlapped)
Modeling facts, statements and lies

Ongoing work with Ludivine Duroyon (U. Rennes 1) and Francois Goasdoué

Approach:
1. **Facts** = RDF
2. **Timed facts** = bitemporal RDF (transaction time, validity time) [Gutierrez, Hurtado 2005]
3. **Beliefs (viewpoints)**: A states that B states that… [Gatterbauer, Suciu 2009]
4. **Timed beliefs**: at time T1, A states that at time T2, B …

Defined:
- Saturation (set of all consequences of a set of statements)
- Query answering

TBD: query answering algorithm
Improving access to reference data

Ongoing work with Tien-Duc Cao and Xavier Tainnier [Semantic Big Data Workshop 2017]

Linked Open Data extraction from INSEE statistic spreadsheets

<table>
<thead>
<tr>
<th>l</th>
<th>Region</th>
<th>Department</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The data reflects children born alive in 2015...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Region</td>
<td>Department</td>
<td>16-20</td>
<td>21-25</td>
<td>26-30</td>
<td>31-35</td>
<td>36-40</td>
<td>41-45</td>
<td>46-50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>Age below 30</td>
<td>Age above 31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Île-de-France</td>
<td>Essonne</td>
<td>215</td>
<td>1230</td>
<td>5643</td>
<td>4320</td>
<td>3120</td>
<td>1514</td>
<td>673</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>Val-de-Marne</td>
<td>175</td>
<td>987</td>
<td>4325</td>
<td>3156</td>
<td>2989</td>
<td>1740</td>
<td>566</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>Ain</td>
<td>76</td>
<td>1103</td>
<td>3677</td>
<td>2897</td>
<td>1976</td>
<td>1464</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Rhône-Alpes</td>
<td>Ardèche</td>
<td>45</td>
<td>954</td>
<td>2865</td>
<td>2761</td>
<td>1752</td>
<td>1653</td>
<td>523</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Poster
Merci / questions?