Fact-Checking the Web: Content Management Technologies for a Democratic Decision Process

Part 2

Ioana Manolescu, OAK
Inria Saclay and U. Paris Sud

Joint work with
François Goasdoué, Konstantinos Karanasos, Yannis Katsis,
Julien Leblay and Stamatis Zampetakis
FactMinder:

A tool based on open standards for fact-checking the Web
XML: extensible markup language

W3C, 2008

clients.xml:

```
<clients>
  <client>
    <nom>Julie</nom>
    <address>1, rue Dugommier</address>
    <city>Paris</city><age>22</age>
  </client>
  <client>
    <nom>Marc</nom>
    <address>2, rue Dugommier</address>
    <city>Orsay</city><age>33</age>
  </client>
</clients>
```

Flexible
Platform-independent
Separate content from presentation
Schema possible (not compulsory)
XML is the main language of Web pages

HTML is fundamentally based on links (which XML preserves). Links to « topic pages » can be seen as a basic primitive to «background information »
XML is the main language of Web pages

XML is the main language of Web pages

HTML is fundamentally based on links (which XML preserves). Links to « topic pages » can be seen as a basic primitive to « background information »
Critique of XML

- Each information ends up in only one place
- OK for "classification" applications, structured text
- Fundamentally restrictive for data = real world!

Tim Berners-Lee, WWW proposal, CERN, 1998:
"Many systems are organised hierarchically. A tree has the practical advantage of giving every node a unique name. However, it does not allow the system to model the real world."
(On newsgroups): "Typically, a discussion under one newsgroup will develop into a different topic, at which point it ought to be in a different part of the tree."
Graph data format for the Web: RDF

• Resource Description Format, W3C, 2003
• Resources have properties with values
• URIs (Universal Resource Identifiers) allow identifying things within namespaces
  - Inria:IoanaManolescu, LinkedIn:IoanaManolescu
• Resources, properties, or values may be specified by an URI.
• Properties and values may be constants
RDF reasoning

• RDF allow expressing data and knowledge

• Example:
  – if someone teaches a class
  – then (s)he is a person, is an instructor, belongs to the school giving the class, and works for the university which includes the school

• Reasoning exploits knowledge to infer implicit data
Our proposed data model

\[ \text{XR} = \text{XML} + \text{RDF} \]

A unified data model for documents and annotations (or facts)

1. **XML documents** such that each XML node has a «virtual» URI (does not have to be materialized in the document or elsewhere)

2. **RDF triples** (subject, predicate, object)
   - XML node URIs allowed in RDF like any other URI

XR data instance = XML instance + RDF instance

\[
\langle \text{Alice}, \langle \text{likes}, \#12 \rangle, \\
\langle \text{Alice}, \langle \text{knows}, _{-}\text{B} \rangle, \\
\#3, \text{rdf:type}, \langle \text{MobilePhone} \rangle \rangle, \\
\langle \#3, \text{rdf:type}, \langle \text{ElectronicDevice} \rangle \rangle \rangle,
\]

\[
\langle \#12, \langle \text{describes}, \#3 \rangle, \\
\#10, \langle \text{date}, \text{"23 May 2011\"} \rangle, \\
\langle \#3, \langle \text{email}, \text{"bob@example.com\"} \rangle, \\
\langle \langle \text{MobilePhone}, \text{rdf:subClassOf}, \langle \text{ElectronicDevice} \rangle \rangle, \\
\langle \#3, \text{rdf:type}, \langle \text{ElectronicDevice} \rangle \rangle \rangle,$
Architecture: the FactMinder prototype

F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu and S. Zampetakis

“Fact-checking and analyzing the Web” (demo)
SIGMOD 2013
FactMinder: fact-checking based on XR
Sample FactMinder view

Bill Clinton apporte son soutien à Barack Obama

Concepts
- Bill Clinton
- Barack Obama
- Des Moines, Iowa
- Hillary Clinton
- Michelle Obama
- Bruce Springsteen
- Pennsylvanie

Facts & figures
- Curriculum
  - Born
  - Political party
  - Spouse(s)

Quotes
- Launching the Africa Regional Media Hub in Johannesburg
- President Obama speaking LIVE for the last time before the election
- Daily Press Briefing: November 5, 2012

Related stories
- Ohio, Floride, ... être compliqué
- Des indicateurs ... d'Obama
- Romney... agressifs
Sample FactMinder view

Rich browser Annotated document

Bill Clinton apporte son soutien à Barack Obama

Concepts
- Bill Clinton
- Barack Obama
- Des Moines, Iowa
- Hillary Clinton
- Michelle Obama
- Bruce Springsteen
- Pennsylvanie

Facts & figures
- Curriculum
  - Born: William Jefferson
  - August 19, 1946
  - Hope, Arkansas, U.S.
  - Political party: Democratic Party
  - Spouse(s): Hillary Rodham

Quotes
- President Obama speaking LIVE for the last time before the election http://bit.ly/PSfky

Sources

5/30/13 Ioana Manolescu / WOD 2013
Sample FactMinder view

«Politician» instantiated views
Sample FactMinder view
Three thoughts for Canadians as Barack Obama bids adieu to war on terror

Canadian policymakers can learn a great deal from the substantive, intelligent foreign policy speech Barack Obama delivered last week.

By Natalie' Bendier
Published on Mon, May 27, 2013

And now, a public service announcement: While Canadians have been gripped by scandal in Ottawa and Toronto for the past two weeks, the global war on terror has ended.

In a major speech last Friday, President Barack Obama declared that after 11 years of wartime footing, a new approach is required. It is time, he said, for the U.S. "to define [its] effort not as a boundless, 'global war on terror,' but rather as a series of persistent, targeted efforts to dismantle specific networks of violent extremists that threaten America."

At the core of Obama’s speech was a resolve to correct pernicious drifts affecting his country. Since 9/11, he noted, its fight against terrorism has succeeded in evening mass terror attacks on U.S. soil — but at a very steep cost. On the domestic side, those costs include nearly a trillion dollars of counterterrorism efforts and a reduced investment in national well-being: over 7,000 Americans dead and many more maligned; and disturbing infringements of democratic freedoms and the rule of law.
Three thoughts for Canadians as Barack Obamabids adieu to war on terror

Canadian politicians can learn a great deal from the substantive, intelligent foreign policy speech Barack Obama delivered last week.

Text size ▼ Font □ Report an Error □ Show story □

LARRY DOWNING / REUTERS

U.S. President Barack Obama makes a point about his administration's counter-terrorism policy at the National Defense University at Fort McNair in Washington, May 22, 2013.

And now, a public service announcement: while Canadians have been gripped by scandal in Ottawa and Toronto the past two weeks, the global war on terror has ended.

In a major speech Friday, President Barack Obama declared that after 12 years of wartime footing, a new approach is required. It is time, he said, for the U.S. to define its effort not as a boundless ‘global war on terror’, but rather as a series of persistent, targeted efforts to dismantle specific networks of violent extremists that threaten America.

At the core of Obama’s speech was a resolve to correct pernicious drifts affecting the country. Since 9/11, however, its fight against terrorism has succeeded in averting mass terror attacks on U.S. soil— but at a very steep cost. On the domestic side, those costs include nearly a trillion dollars of counterterrorism spending that raised deficits and foreclosed investment in national well-being; over 7,000 Americans dead and many more maimed; and disturbing infringements of democratic freedoms and the rule of law.

On the international side also, the costs to America’s interests have been...
Creating People view using XRQL
Overview of People instances in the Web page

Three thoughts for Canadians as Barack Obamabids adieu to waron terror

Canadian politicians can learn a great deal from the substantive, intelligent foreign policy speech Barack Obama delivered last week.

By: Larry Downing / Reuters

U.S. President Barack Obama makes a point about his administration’s counter-terrorism policy at the National Defense University at Fort McNair in Washington, May 23, 2013.

And now, a public service announcement: while Canadians have been gripped by scandal in Ottawa and Toronto for the past two weeks, the global war on terror has ended.

In a major speech Friday, President Barack Obama declared that after 10 years of wartime footing, a new approach is required. It is time, heralded, for the U.S. “to define its effort not as a boundless ‘global war on terror,’ but rather as a series of persistent, targeted efforts to dismantle specific networks of violent extremists that threaten America.”

At the core of Obama’s speech was a resolve to correct perceived drifts affecting his country. Since 9/11, he noted, “its fight against terrorism has succeeded in averting mass terror attacks on U.S. soil – but at a very steep cost.” On the domestic side, those costs include nearly a trillion dollars of counterterrorism spending that raised deficits and forestoded investment in national well-being; over 7,000 Americans dead and many more maimed; and disturbing infringements of democratic freedoms and the rule of law.

On the international side also, the costs to America’s interests have been staggering. Its reputation has been deeply damaged through drone strikes on civilians and extra-judicial detentions at Guantanamo Bay, and it has failed to help resolve major threats to the world like the enduring entrenchment of regime after regime and the threat from climate change.

To read more about this story, please visit thestar.com.
Interaction between XIPs

(1) Barack Obama is selected

(2) Dependent XIPs are updated to present the new content
Three thoughts for Canadians as Barack Obamabids adieu to war on terror

Canadian politicians can learn a great deal from the substantive, intelligent foreign policy speeches Barack Obama delivered last week.

By Natalie Bennett Published on Mon May 27 2013

And now, a public service announcement: while Canadians have been gripped by scandal in Ottawa and Toronto for the past two weeks, the global war on terror has ended. In a major speech late Friday, President Barack Obama declared that after six years of wartime footing, a new approach is required. It is time, he said, for the U.S. "to define [its] effort not as a boundless 'global war on terror,' but rather as a series of persistent, targeted efforts to dismantle specific networks of violent extremists that threaten America."

At the core of Obama's speech was a resolve to correct pernicious drifts afflicting the country. Since 9/11, he noted, "the fight against terrorism has succeeded in averting mass terror attacks on U.S. soil — but at a very steep cost. On the domestic side, those costs include nearly $2 trillion dollars of counterterrorism spending that raised deficits and foreclosed investment in national well-being over 7,000 Americans dead and many more maimed and disturbing infringements of democratic freedoms and the rule of law."

On the international side also, the costs to America's interests have been staggering. Its reputation has been deepened through drone strikes on civilians and extra-judicial detentions at Guantanamo Bay, and it has failed to help

---

Interaction between XIPs
Annotating new content with FactMinder

1. Select the part of the web page to be annotated (highlighted with yellow).
2. Specify the ontology which will be used for the annotation.
3. Specify the type of the selected resource.
4. Add a comment.
5. Save the annotation.
Annotating new content with FactMinder

(1) Select the part of the web page to be annotated (highlighted with yellow)

(2) Specify the ontology which will be used for the annotation

(3) Specify the type of the selected resource

(4) Add a comment

(5) Save the annotation

Two triples are generated and saved to the XR server:

- Triple(http://www.thestar.com/.../#uniqueuri1, rdf:type, oce:Person)
- Triple(http://www.thestar.com/.../#uniqueuri1, rdfs:comment, “Natalie Brender is probably a writer for thestar.com web page”)

5/30/13
Ioana Manolescu / WOD 2013
Another annotation example

The previous annotation is saved

New annotation
XR model, language and algorithms
FactMinder views = XR queries

Simple XR queries: head is just tuples of bindings

Return people who like content describing electronic products, and the contact information for these products manufacturers
Extended XR query language

Finished formalization with Alin Deutsch (UCSD) Currently working on query-view composition

Query result is “new” XR instance

Construct a set of trees, gathering information about popular products, with annotations showing people who advocated these products, and references to the content.
Evaluating XR queries

Relying on an XML Data Management platform (**XDM**) and an RDF Data Management platform (**RDM**)

Focus on simple queries (no XR result construction)

To do:
- Eval \( Q_X \)
- Eval \( Q_R \)
- Join the results
XR query evaluation

• How to handle implicit XML URIs?
  – This is what queries join on
  – But the URIs are not in the data...

• Two independent hypothesis on the XDM
  – XURI-in: the XDM can select on a given XML node ID
  – XURI-out: the XDM returns XURIs in its output

• None, one, or both may hold

• The XR engine adapts to the existant
XR query evaluation considerations

• What to do if XURI-in does not hold?
  – **Dereference** XURI into linear parent-child path
  – Push the parent-child path in the XML sub-query and have it evaluated

• How many queries to send to the XDM?
  – One per tuple of bindings (« tuple-at-a-time »)
  – A single one (« set-at-a-time »)

• Push-to-data option: materialize bindings of the sub-query on data model 1, into data model 2
XR query evaluation strategies

- Independent execution
  - XML first*
    - Push to data
      - XML → RDF-Data
      - RDF → XML-Data
      - RDF → XML-Data-Pr**
  - RDF first
    - Push to query
      - Many queries
        - XML-in holds
        - RDF → XML-URI
        - RDF → XML-URI-Pr**
      - Single query
        - XML-in does not hold
        - Many queries
        - RDF → XML-XPath
        - RDF → XML-XPath-Pr
      - Single query
        - RDF → XML-XPath-Pr

- Dependent execution
  - Push to data
    - XML → RDF
  - Push to query
    - Many queries
      - XML-in holds
      - RDF → XML-URI
      - RDF → XML-URI-Pr**
  - Single query
    - XML-in does not hold
      - Many queries
        - RDF → XML-XPath
        - RDF → XML-XPath-Pr
      - Single query
        - RDF → XML-XPath-Pr

* XURI-out holds
** Uses dereferencing for efficiency
Strategy **XML→RDF**

Evaluating $Q_X$ binds the variables $A$ and $B$  
For each $(A, B)$ binding pair, instantiate $Q_R$ and send it to the DM
Strategy RDF $\rightarrow$ XML-URI (assuming XURI-in)

Evaluating $Q_R$ binds the variables $\$A$ and $\$B$

Enrich $Q_X$ with the conditions

« and xuri($v1)$=$\$A$ and xuri($v2)$=$\$B$ »

Send it to the XDM
Strategy evaluation

- Implementation on RDF-3X and (BaseX | ViP2P)
- If XURI-out, XML | | RDF and XML → RDF win and scale well
- On small databases, if XURI-in, use RDF → XML-URI; otherwise RDF → XML-XPath-Pruning
Related works

• Standards and tools for annotated documents
  – We give first-class citizenship to XML and RDF

• Converting RDF to XML or viceversa?
  – We believe both formats are very helpful as they are: XML for structured serialized text, RDF for atomic fcts

• XML-RDF interoperability
  – We focus specifically on making XML nodes referenceable in an RDF data set
Conclusion / discussion
Fact-checking is incredibly useful

• Basic step of any scientific effort.
• But we’re not all scientists.
• Are 5500 facts enough? (Maybe?)
• How to persist, interconnect and exploit separated databases of checked facts?
  – French press still in text mode (no metadata)
  – Standard-based solutions like XR should help
• Incentives/disincentives? Google’s blind angle?
Discussion

• Who owns this?
  – Maybe we should take up a class on Organization Theory 101

• Who can federate this? Wikipedia? EU project, necessarily with social scientists or journalists?
  – Interaction with LOD projects themselves...
  – US vs. Europe

• What about social networks?
Merci / questions / more discussion?