Contextualization of Facebook documents

Diane Gallois-Wong, Raphaël Rieu-Helft

February 22, 2016

Facebook documents

- **Photograph** : a lot of information in metadata
- Facebook event : rich context to study the photographs

The Neemi platform

- Django webapp that extracts user data from various social networks and stores it in a MongoDB database
- Our project is a fork of Neemi with added features

Contextualizing a Facebook document

- Choose a photo or event in the database
- Build a corresponding RDF graph
- Fetch related items in the database
- Build their RDF graphs and absorb them into the first one

Facebook API limitations

- The latest versions of the API is very restrictive with data from friends who didn't install our app
- Getting interesting data is difficult
- New Neemi feature : add mock documents

Absorbing graphs : algorithm

Base graph A representing an event
Absorb a graph B representing an event or a photograph

- Copy triples of B into A : nodes of B need representatives in A
- Literals, URIs : representative is self
- Instances : if node for the same unique thing already in A (e.g. people) then this, else new blank node
- Infer relations : people on a photograph attended the corresponding event...

Some technical choices

- Convert each photograph and event to an RDF graph, only to immediately absorb it into a main graph : more complicated here, but more general
- Applying time of photograph creation to a corresponding event : "startBefore" and "endAfter" properties
- Deciding if a document is relevant to the event : time-based and quite lenient for now, can be improved

Potential extensions

- Handle documents from other sources than Facebook : new methods need to be added but no conceptual difficulty
- Better heuristics to deduce if an item is related to the event