CHAPTER 8
SUCCESS STORIES

Martin Liahagen
Agenda

• A Knowledge Graph for Innovation in the Media Industry
• Applying Knowledge Graphs in Cultural Heritage
• Applying Knowledge Graphs in Healthcare
A KNOWLEDGE GRAPH FOR INNOVATION IN THE MEDIA INDUSTRY

HAVAS 18 Innovation Labs
The Business Problem

- The way consumers and brands communicate is changing dramatically
  - Personalisation
  - New communication channels
  - New technology

- HAVAS 18 Innovation Labs
  - A “better together” approach
  - Wants to connect innovators, universities, start-ups and technology trends
The HAVAS 18 Knowledge Graph I

- “Enable search, discovery and understanding of information about start-ups in their first 18 months”

Data sources:
- Generalist and specialist web sites for core data
  - AngelList, CrunchBase
- Social networks for relationship data
  - Facebook, LinkedIn, Twitter
- Newsfeeds for extended media coverage
The HAVAS 18 Knowledge Graph II

■ Structured and integrated in RDF
■ Schema built on top of
  - Schema.org
  - FOAF
  - SKOS
  - rNews
■ RESTful API with JSON
  - API services include CRUD methods

■ Rapporteurs
  - Can add and modify entities and relationships
  - Can also be curators
Value Proposition

- Gathers insight and information automatically
  - *Meaning business strategists have something to start with*
- Can be used to evaluate trends and start-ups
- Provides knowledge about the best-in-class talent for new technology
Challenges

- Entity resolution and disambiguation
  - *Evidence model, with key classes*
    - E.g. start-ups could have Founder, Client and Technology
    - *Knowledge tagger extracts entities, which are matched with evidence models*

- Version management
- Resilience against changes in data sources, like web APIs
- Monitoring potential decay
APPLYING KNOWLEDGE GRAPHS IN CULTURAL HERITAGE

The CURIOS Project
Digital Cultural Heritage

- Projects
  - UK Culture Grid
  - Continuous Access to Cultural Heritage (CATCH)

- Ontologies
  - Categories for the Description of Works of Art
  - International Committee for Documentation Conceptual Reference Model (CIDOC CRM)

- Technologies
  - XML
  - Distributed databases
  - RDF/OWL
Digital Cultural Heritage and Linked Data

- OpenART
  - *The London Art World 1660 – 1735*
  - *Linked Open Data format*
- CultureSampo
  - *Cultural heritage archive for Finland*
- datos.bne.es
  - *Library data from The National Library of Spain*
Challenges

- Data heterogeneity
  - Data formats are not consistent
  - Makes it hard to integrate different archives

- Computer literacy of the community
  - Ranges from novice to experienced
  - Current tools require in-depth technical knowledge
The CURIOS Project I

- CUlтурal Repositories and InfOrmation System
- Solves both challenges
- Data heterogeneity
  - Uses Linked Open Data as standard
- Computer Literacy
  - Uses Drupal, a Content Management System (CMS)
  - Does not require in-depth knowledge
The CURIOS Project II

- Loosely coupled
  - Data is stored in RDF
  - Data presentation configurations stored in Drupal’s SQL database

- Supports semantic searching via SPARQL, configurable presentation and visualisation services.

- Used for
  - preserving cultural heritage
  - improving tourist experiences and exhibitions
  - storing research data
Constructing the Knowledge Graph

Domain ontology
- Extended from a simple upper ontology
- Basic properties inherited from the upper ontology for special presentation (images, dates, metadata)

Meta-ontology
- Configure presentation
- Define mappings between domain ontology’s classes and properties to CMS’s content types and fields
- Allow validation of presentation

CMS’s entities
- CMS’s entities and fields are stored in the CMS’s database while concrete data (RDF triples) are stored in a triple store.
- Use CMS’s entities to build webpages’ templates for human users to browse

Mapping from a domain ontology to CMS’s entities
Presentation of Instances

4 Caverstay

Croft 4 was first occupied by Donald Mackinnon and then by his son Roderick.

The croft was particularly congested in the early years of the 20th century, supporting four large families of Mackinnons. The situation forced many to leave the village for Stornoway, the mainland or abroad.

<table>
<thead>
<tr>
<th>Title</th>
<th>4 Caverstay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record Type</td>
<td>Crofts and Residences</td>
</tr>
<tr>
<td>Gaelic Name</td>
<td>4 Cladh-choille</td>
</tr>
<tr>
<td>Type</td>
<td>Croft</td>
</tr>
<tr>
<td>Record Owned By</td>
<td>CEP</td>
</tr>
<tr>
<td>Record Maintained By</td>
<td>CEP</td>
</tr>
<tr>
<td>Subject Id</td>
<td>7960</td>
</tr>
</tbody>
</table>

Lived Here

- Angus Mackinnon
- Johanna Mackinnon
- Donald Mackinnon
- Mary Bell Macleod
- Ann Mackinnon
- Catherine Macleod
- Roderick Macleod
- Donald Mackinnon
- Louis Mackinnon
- John Mackinnon
- Mary Ann Mackinnon
- Louis Mackinnon
- Mary Macleod
- John Murdo Macleod
- Catherine Macleod
- Christina Mackinnon

Associated With

- Euphemia Maciver
- Ruairidh Rob Mackinnon: Memories...
- Ruairidh Rob Mackinnon: Off to...
- John Murdo Macleod
- Catherine Macleod

Located At

Caverstay
Dealing with Vagueness

- Inexact dates and times
- Fixed by using ranges
  - E.g. 1780 becomes [1780-01-01, 1780-12-31]

<table>
<thead>
<tr>
<th>General Class</th>
<th>Pattern</th>
<th>Example</th>
<th>Frequency</th>
<th>Subtotal</th>
<th>Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact to the day</td>
<td>y-m-d</td>
<td>1780-06-13</td>
<td>12949</td>
<td>13954</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td>d-m-y</td>
<td>10/6/45</td>
<td>725</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d-M-y</td>
<td>12 MAY 1780</td>
<td>272</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M-d-y</td>
<td>May 12 1780</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact to the month</td>
<td>y-m</td>
<td>1780-12</td>
<td>274</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M-y</td>
<td>Aug 1780</td>
<td>443</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m-y</td>
<td>03/1780</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact to the year</td>
<td>y</td>
<td>1978</td>
<td>10825</td>
<td>10825</td>
<td>yes</td>
</tr>
<tr>
<td>Exact to the decade</td>
<td>dec</td>
<td>IN 1860's</td>
<td>1415</td>
<td>1415</td>
<td>yes</td>
</tr>
<tr>
<td>Exact to a range of years</td>
<td>y-y</td>
<td>1939-45</td>
<td>242</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>before</td>
<td>pre 1918</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>after</td>
<td>AFT 1890</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact to the century</td>
<td>cent</td>
<td>20th Century</td>
<td>4</td>
<td>4</td>
<td>yes</td>
</tr>
<tr>
<td>Vague within less than a month</td>
<td>mend</td>
<td>Aug/Sept 1972</td>
<td>26</td>
<td>26</td>
<td>yes (using a date range)</td>
</tr>
<tr>
<td>Vague within more than a month but less than a year</td>
<td>y-end</td>
<td>1978/70</td>
<td>7</td>
<td>7</td>
<td>yes (using a date range)</td>
</tr>
<tr>
<td>Vague year</td>
<td>cy</td>
<td>C. 1932</td>
<td>566</td>
<td>652</td>
<td>yes (using a date range)</td>
</tr>
<tr>
<td></td>
<td>moddec</td>
<td>early 1950s</td>
<td>86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vague around a decade</td>
<td>cdec</td>
<td>c 1950s</td>
<td>2</td>
<td></td>
<td>yes (using a date range)</td>
</tr>
<tr>
<td></td>
<td>modcent</td>
<td>LATE 1600S</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not directly interpretable as a date</td>
<td>rank</td>
<td>D.I.I.</td>
<td>3069</td>
<td>3069</td>
<td>no</td>
</tr>
<tr>
<td><strong>GRAND TOTAL</strong></td>
<td></td>
<td></td>
<td>30923</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of date forms
APPLYING KNOWLEDGE
GRAPHS IN HEALTHCARE

IBM
Clinical Practice Guidelines (CPGs)

- “Collective sets of treatment recommendations that attempt to capture the best medical practices for different pathologies”
- They become more complex as new medical treatments are discovered
- CPGs recommend Treatment Programmes (TP)
  - TP is a sequence of medical procedures
- Deviations from the CPGs are not wanted, but...
- CPGs cannot cover everything
Diagnose the patient's clinical presentation

Decide whether to prescribe one of the CPG recommended TPs for the diagnosed clinical presentation

Prescribe a TP according to CPG recommendations

Prescribe a TP discordantly to CPG recommendations

Manage the actual TP that is performed according to CPG recommendations

Manage the actual TP that is performed discordantly to CPG recommendations

No deviation

Deviation
The Project

- **Goal:** Automatically monitor compliance to CPGs and identify potential rationale for deviations
- **How:** Computationally parse Electronic Health Records (EHRs) using
  - *Natural Language Processing*
  - *Data modelling*
  - *Comparison algorithms*
- Data based on EHRs of patients in Italy with Soft-Tissue Sarcoma
  - *Rare cancer with numerous treatment options*
Natural Language Processing Techniques

- Italian to English machine translation
- Processed the unstructured information with UIMA framework
  - Part-of-speech tagging
  - Normalisation
  - Entity and relationship extraction
  - Semantic analysis
  - Negation
  - Disambiguation reasoning
Services Based on the Knowledge Graph I

- Treatment programme comparison
  - Compare the actual and the recommended TP
  - Decide how much the actual TP deviates

- Classifications of deviations
  - Over- or undertreatment
  - Difference in chemotherapy drug
Services Based on the Knowledge Graph II

- Extracting justifications
  - “In light of extension of illness, the patient’s age and preliminary activity of molecule in this particular histotype, starting chemotherapy with gemcitabine.”

- Classification of reasons
  - Cancer status, previous treatments, patient preference
Contributions

- Helping doctors understand clinical deviations
- Helping to understand the decision-making process of physicians
- Improving CPGs