Web & Media Group
MSc Projects

Lora Aroyo

VU UNIVERSITY AMSTERDAM
We work with ...

- Social Sciences
- Communication Sciences
- Languages
- History

VU University Amsterdam

• synerscope
• ELSEVIER
• Koninklijke Bibliotheek
• IBM

Rijksmuseum Amsterdam

• DBpedia
• MediaMatic

BBC

NCB Naturalis

• Kro
• DANS

Philips

Above Water Systems

Logica CMC

Waag Society

WIKIDATA

Thales
MSc Projects Themes

Data, Analysis & Visualization
Cultural Heritage, Web & Visitors
Future TV & Web & Social Web
Collection & Metadata Enrichment
Interactive Mobile & Web Apps
IF YOU LIKE ...

• experimenting
• exploring & analytics
• to be driven by curiosity
• pragmatics-oriented approaches
• creative working spirit
• having fun
• general 'hacker' attitude
PROJECTS:
CULTURAL HERITAGE, THEIR VISITORS & THE WEB
Cultural Heritage, Their Visitors & the Web

- Data Enrichment
- Personalized semantic search
- Mobile Museum & City Tours
- Interactive Multitouch Applications
- Personalized access to museum collections
- Games with a purpose
- Crowdsourcing for Video & Image Tagging
- and many more ....
INTERACTIVE, ADAPTIVE, SEMANTIC MUSEUM TOURS

http://www.chip-project.org/demo/
INTERACTIVE, ADAPTIVE, SEMANTIC CITY TOURS

Thursday, October 3, 13
PERSONALIZED SEMANTIC SEARCH

http://sealincmedia.wordpress.com/
http://semanticweb.cs.vu.nl/clustersearch
PERSONALIZED SEMANTIC SEARCH

http://sealincmedia.wordpress.com/
http://semanticweb.cs.vu.nl/clustersearch
PERSONALIZED SEMANTIC SEARCH

http://europeana.eu

http://e-culture.multimedian.nl
CROWDSOURCING

IMAGE TAGGING

Veuë de Marcoussy du costé de Montleberty
Gezicht op een kasteel met een roeiboot in de slotgracht. Op de voorgrond een bosachtig landschap met ruiters te paard.
Helping Humanities Scholars to Search the Rijksmuseum Prints Online Collection

Humanities researchers depend in their research on the efficiency and effectiveness of the search functionality provided in various cultural heritage collections online (e.g. images, videos and textual material). Currently many of the cultural heritage institutions do not provide the necessary interactivity and transparency intuitive for humanities scholars. The Rijksmuseum collection is an important research source for many humanities researchers worldwide. One of the important subcollections is the Prints Cabinet Online (containing more than 600,000 number of prints in various genres). With the current search implementation humanities scholars can primarily search and collect information about individual objects. Exploratory research indicates that humanities scholars (i.e. art historians, cultural historians, book historians, historians of science, of religion of literature, etc.) create complex, but unarticulated ways to search online cultural heritage collections, and their needs for analysis and deeper reflection over clusters of artefacts and concepts is still insufficiently supported. The main goal of this project is to study the user needs of humanities scholars and implement them in a demonstrator, which will serve as a basis for an evaluation pilot to collect user feedback on a novel semantic search approach.

Our approach proposes clustering of search results based on semantic patterns in linked cultural heritage data. Through the feedback from the user studies, proposed in this project, we will be able to rank the semantic patterns by their importance for each specific arthistorical genre. We believe that, in this way, we will ultimately help in providing the necessary functionality for the formulation, refining and answering of humanities research questions. Thus, the main hypothesis of this research is that we can define genrespecific relevance values for linked data patterns that would support not only better semantic search, but also the analysis process of humanities scholars.
PROJECTS:
DATA ENRICHMENT
Enriching Museum Collection Metadata

Lot and his Daughters

A seasoned but lusty old man is seated between two naked young women. In the background a city is burning. The man is Lot, seduced by his daughters following the destruction of the city of Sodom. Hendrik Goltzius painted the work in 1616. He used the Bible story to show off his skill as a painter of nudes. The two women have wonderfully soft bodies with full, gentle curves. For an old man, Lot is still remarkably muscular. To accentuate the bodies the artist draped cloth over them in contrasting colours: blue, yellow and red. The poses of Lot and his daughters are perhaps rather artificial, but that was the style of art in the period, Mannerism.

Title
Lot and his Daughters

Year
1616

Technique
Oil on canvas

Dimensions
140 x 204 cm

Object number
SK-A-4866
ENRICHING MUSEUM COLLECTION METADATA

Thursday, October 3, 13
ENRICHING MUSEUM COLLECTIONS
Enriching Museum Collections
Enriching Museum Collections

Tour de Rijks
Rate these artworks

River View near Deventer
Creator: Ruysdael, Salomon van

Why River View near Deventer is given as a recommendation?

Artwork has the following properties that you rated positively:
- Oil paint
- Unprimed canvas
- Dutch landscapes

Artwork has the following properties that were recommended to you:
- Haarlem
- Water, ice and snow
- Netherlands and the Water

View related artworks

Thursday, October 3, 13
Projects: Data Visualization & Analysis
Visualization

Investigate the relationship between physical interfaces/environments, visualization and web standards: Javascript (d3.js), PROV

Source: Department of Energy & Climate Change, Tom Counsell.

Paul Groth (p.t.groth@vu.nl)
Community Analysis

Understanding specific communities on the web.

For example, how does the web community develop standards?

Paul Groth (p.t.groth@vu.nl)
Understanding scientific spreadsheets (I)

Goal
Method of (semi)automatic spreadsheets analysis, to understand included concepts and relations

Case study
Microtax model

CPB Netherlands Bureau of Economic Policy Analysis
Calculates purchasing power Dutch population

Freely available, but difficult to understand
Understanding scientific spreadsheets (II)

Activities
• Analyzis of table architecture
• Semantic characterization of terms
• Reconstruction of underlying ontology

Possible extensions
• Visualize results
• Analyze calculation workflow
• Link spreadsheets with related publications
• Convert spreadsheets into linked data

Martine.de.Vos@vu.nl
Projects: Future TV

NoTube: Personalized, Social & Interactive TV

ViSTA-TV: Linked Open Data, Live Analytics, Recommendations
INTERACTIVE TV

SHARED & PERSONALIZED SECOND SCREENS

http://www.notube.tv
The Trailer Maker

This service builds automatically video trailers. Currently the service is under testing, click here to contribute to the evaluation!!

Infinite Trailers

This app is a new EPG concept where the user sees short trailers of TV programs instead of a list, when looking for a program to watch. We are currently testing this app, click here to contribute to the evaluation!!
ViSTA-TV project

Project proposals:

- Visualize Open Web recommendations
- Social networks and recommendations

SIRUP: Enhancing Serendipity In Recommendations via User Perceptions

Creating serendipity (i.e. “pleasant surprises for users”) is a primary goal of intelligent recommender systems. This project proposes an interdisciplinary approach to enhance the serendipity of TV recommendations that combines complementary knowledge from three disciplines – Computer Science, Language & Cognition and Communication Science. The project examines the “back-end” or algorithms behind serendipitous TV recommendations (Computer Science), the “front-end” or the actual display of these recommendations (Language & Cognition), and the “effect” on users’ perceptions and satisfaction (Communication Science). The project builds on an existing recommender system that combines (a) enriched content semantics with linked data and (b) broadcaster viewing data, developed by one of the applicants in the context of ViSTATV (http://vista-tv.eu). We seek to further enhance this system by adding insights about serendipity both from the perspective of Language & Cognition and Communication Science.

Scenario: Current TV recommender systems typically rely on genre or actors, e.g., recommendations referring to other “suspenseful genres” or “documentaries”. However, less obvious relationships between programs (e.g., relationships between actors and directors, or settings and locations) exist that can be found in linked data and exploited for more serendipitous recommendations. However, to date, little is known how these “linked data patterns” can be exploited to maximize the serendipity of TV recommendations. Accordingly, the present project seeks to fill this research gap. More specifically, we ask:
Project: Web Apps & Social Web
The Social Web

- **Crowdsourcing**: video annotation games
- **Hacking History**:
  - social portal for the interested in history
  - finding event-related clusters of multimedia
- **Internet Barometers - Online Activism**
  social network data analysis
- **Linking Twitter to other sources**
Winner EurolITV Competition
Best Archives on the Web Award

http://waisda.nl
Linking Tweets to other sources

- Can we link Tweets to the political events, newspaper articles and TV shows that they talk about?
- Can we automate an analysis of which news is tweeted about most?
- Can we link tweets of politicians to what they say in parliament or in the media?
Comparing Wikipedias

- Analyze the differences between the content of the Wikipedias in different languages. For example:
  - What entities have a page on the Dutch Wikipedia, but not on the English one?
  - Do countries that are geographically near each other cover more of each others people, places, events, music, etc?
• Can algorithms mimic human trust?

• What are the characteristics of trustworthy pieces of text?

• Can provenance help determine the trustworthiness of artifacts?
Extractivism: Extracting & Visualising Activist Events

**Goal**
Support organisational scientists studying activist organisations through extracting, aggregating, and visualising the events they participate in.

**Why**
Activists could have a significant role in shaping social views and opinions, e.g., Occupy, Arab Spring.

**Problem**
Incomplete, incorrect, uncontextualized, and biased information.

**Solution**
Extract event mentions from different news sources that could complement, contradict, or verify each other. Visualize the aggregated results to support analysis and discovery.

**Who**
Computer Science: Thomas Ploege, Chun Fei Lung, Lora Aroyo
Organization Sciences: Maxine Krujff, Frank de Bakker, Iina Hellsten
www.mona-project.org

---

**Extraction Pipeline**

Our pipeline extracts events from news articles using NLP technology such as Stanford CoreNLP and TextRazor’s API.

1. We split the article into sentences and words.

2. We identify verbs - these are the Events.

3. Any named entities become Actors, Places, or Times depending on their type.

4. We parse grammatical dependencies to relate Events to Actors, Places, and Times.

5. Together, the verb and its dependent named entities form an event as shown in the graph above. We use the Simple Event Model (SEM) as a schema.

6. The process is repeated for every article and sentence in each article. The resulting events are exported as RDF or JSON for further usage in visual analytics applications.

**Visual Analytics**

Geographical overviews help answer questions like: What is the intensity of activist activity in certain areas? What is the chronological spread from one area to another?

Event type clouds help answer questions like: What is the current state of a conflict that activists are involved in? Are there primarily things being said, or are more concrete actions undertaken?

lora.aroyo@vu.nl
http://crowd-watson.nl/dr-detective-game/

Dr. Detective

combining gamification techniques and crowdsourcing to create a gold standard in medical text

Step 1: Select the type of clue you are looking for.

Step 2: To pick a clue, highlight all the words that describe it by clicking on them.

Step 3: After all the words in the clue are highlighted, save the clue.

Step 4: After you found all the clues for Factors, submit them.

Step 5: Go back to step 1 and select another clue type, or move on to the next diagnosis.

In the following text, find all the clues that could help diagnose malignant melanoma, invasive to 0.9 mm, with metastasis to one of two sentinel lymph nodes.

On physical examination, the patient appeared healthy, and the vital signs were normal. There was a 3-mm linear scar on the left upper posterior arm, with no residual pigmented lesion and no satellite nodules. No other suspicious pigmented lesions were identified. There was no palpable lymphadenopathy. The remainder of the examination was normal.
Analytics for Crowsourced Text Annotations

http://crowd-watson.nl
Analytics for Crowsourced Text Annotations

http://crowd-watson.nl
Projects:
More on the Web & Cultural Heritage
Projects: Interactive Interfaces for Smart Phones, Tablets, Multitouch Tables ...
INTERTAIN@VU
EXPERIMENTATION

http://www.networkinstitute.org/tech-labs/intertain-lab/
NO NEED TO FIGHT,
THERE ARE ENOUGH PROJECTS FOR EVERYBODY

EMAIL US:
WEB-MEDIA-MASTER-L@FEW.VU.NL
LORA.AROYO@VU.NL