Towards Content-Rich Digital Genealogy with Model-Oriented Approaches:
Issues and Standards

Robert B. ALLEN　ロバート　アレン
Research Center for Knowledge Communities
University of Tsukuba
Tsukuba, Japan

rba@boballen.info

Rich Semantic Descriptions via Modeling

- As we see from the conference, there are a great many facets of interest in digitized newspapers. Can we develop a general framework to support organization, context, and access?

- Rather than developing indexes for specific terms, perhaps we can model complex entities and the what they are doing.

- Modeling would show the network of relationships and could be dynamic to include states, behaviors ,and processes.

- Models could be developed for many types of entities such as people, families, multi-families, communities, and cities.

- Here, we focus on communities and the people in those communities since those match the historical newspapers.

Start with Entity-based Descriptions (e.g., People)

- There are several emerging but limited standards for describing people:

  – FOAF – Friend of a Friend

  – GEDCOM - GEnealogical Data COMmunication

  – People Australia http://peopleaustralia.anu.edu.au/

Awards	Death state	Events	Properties
Birth town	Death country	Gender, Groups	Religion
Birth state	Educational institution (higher)	Military service	Workplaces
Birth country	Educational institution (school)	Occupations	
Cause of death		Organizations	
Death town	Ethnicity	Places	

  – "Interactive directory" for Norfolk, Nebraska

  Allen, R.B., Toward an Interactive Directory for Norfolk, Nebraska: 1899-1900, IFLA Newspaper and Genealogy Section Meeting, Singapore, Aug 2013. arXiv:1308.5395
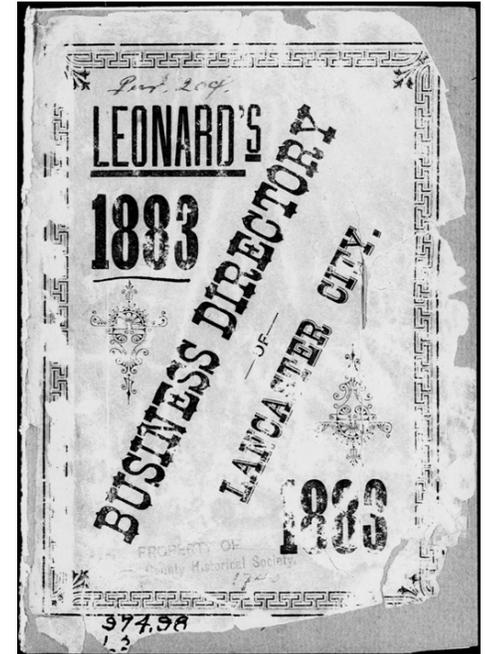
Models that include Behaviors

- We need models which describe behavior, situations, and contexts of entities.
- We have explored using object-oriented models implemented with Java for descriptions of people and communities.
- Object-oriented models are useful because they have:
  - Classes (objects) model entities with attributes as a package
  - Behaviors are implemented with methods which are part of the classes
  - Java encourages simple approaches to abstraction and instantiation
- We can use the models for simple descriptions or we can develop executable models which unfold as they are run.

Allen, R.B., Frame-based Models of Communities and their History. *Histoinformatics*, Nov 2013, Kyoto..

Using FrameNet Frames to Structure Java Classes for Modeling Texts

- We have lots of rich text descriptions from cultural descriptions. Could we use that? After all, <u>the text descriptions are representations.</u>

- One approach to modeling would be to model the text with FrameNet (<u>https://framenet.icsi.berkeley.edu/fndrupal/</u>)

  - Essential concepts in natural language described with frames. These frames are related to semantic roles.

  - FrameNet is based on cognitive principles but we use it as a language resource for modeling.

  - We are particularly interested in verb frames because they describe transitions in attributes.

    - About 700 verb frames.

    - Example: Frame for "Release"

      A Captor ends the captivity or inhibition of the motion of a Theme from the Location_of_confinement. The release is in accord with the plans of the Captor.

Allen, R.B., Model-Oriented Information Organization: Part 1, The Entity-Event Fabric, D-Lib Magazine, July 2013.
Allen, R.B., Model-Oriented Information Organization: Part 2, Discourse Relationships, D-Lib Magazine, July 2013.

Example Text from Wikipedia
"Early history (1750-1820) of Minneapolis, Minnesota"

- Used textbook or Wikipedia-level texts
  - These are relatively straight-forward, with simple past tense
  - By comparison, primary sources have many difficulties. Full of slang, complex constructions, un-grammatical, and often incorrect statements.
  - Some massaging is still required
- Includes a passage about "releasing" captives.
- Verb frames can be modeled via methods in Java class.
- So far we have a "proof of concept" but:
  - Limitations of frames
    - Not always a perfect match
  - Limitations of Java
    - Does not support multiple inheritance

Building Comprehensive
Person Models and Community Models

- While Frames focus on text descriptions, additional information about the community may be more readily available from other sources (census, diaries, oral histories).
- By combining sources of information at different levels, we can develop "community models" and "person models". Many communities are relatively closed systems. This makes them more tractable than indexing, say, cities.
- Community models must also include Person models
  - Explanations based on theories of mental events and human motivation.
  - Interact with community models.
    - Communities structure human a lot of behavior. This structure help us to describe a person's behavior
  - Can show internal dynamics such as;
    - Time course of emotion
    - Maturation across years
- These models need complex structures for entities and processes.
  - Classification (inheritance) hierarchies
  - Partonomies
    - Grouping like-objects
    - Hierarchy of parts
  - Systems and functionality

Discourse Models to Support Presentations and Interaction

- In addition to the content model, we want ways to support interaction and presentations.
- Discourse overlay
  - Explanations
    - Tutorial-like overviews of people and communities
  - Argumentation
    - Structured comparison of generalizations and of evidence
- Supporting history and genealogy researcher
  - Interactive interfaces for working with community histories
- Managing the content model
- Visualizations, Narratives, and Games
  - Interactive historical re-enactors, games, and cyber-dramas

Allen, R.B., Visualization, Causation, and History, *iConference*, 2011.

Status and Standards

- We have:
  - Shown first steps to developing structured models of communities with rich semantics
  - Shown it is possible to model natural language texts with FrameNet. However, there are often gaps and we to develop descriptions that cross many levels.
- Standards:
  - Need to develop additional structures for:
    - Complex entities such as people and organizations
    - FrameNet frames generally works well but they need to be extended.
    - Several aspects of incorporating natural language are not yet well explored.
  - Need community involvement in setting standards

Issues and Future Directions

- Broader effort to develop model-oriented information organization
  - Extend model-oriented information organization to museum objects and informatics
  - The benchmark for semantics is human cognition. To what extent should our models be based in cognitive modeling?
  - Modeling nuances of natural language such as representing the relationship of clustered terms like: Baker – Bakery - Baking
- Extend community models to:
  - Multi-family genealogies
  - Networks of communities
  - Modeling cities, nations

Towards Content-Rich Digital Genealogy with Model-Oriented Approaches:
Issues and Standards

Robert B. ALLEN　ロバート　アレン
Research Center for Knowledge Communities
University of Tsukuba
Tsukuba, Japan

As of March 1, 2014
Robert B. ALLEN　로버트 알렌
Department of Library and Information Science
Yonsei University
Seoul, Korea

rba@boballen.info