Representing knowledge in the semantic web

ISS Proseminar - F15

September 24, 2015

Given a domain of interest to us, and the concepts and knowledge associated with it

- … how do we **describe** the domain of interest?
- … how do we **store** and **retrieve information** about the domain of interest that is relevant?

What does it mean to describe something?

What does it mean to describe something?

- Provide details about it
  - What are those details?
- Entities/objects
- Relationships among those entities

Let’s think about the domain of autobiographical knowledge

- what sort of information constitutes knowledge we have about ourselves and our past?
- why is knowing things about someone’s past useful, say for various companies that do their business on the web?

We need to store knowledge in some sort of database

- The database has to be organized in some way
- What are ways in which we can organize the data in a database?

Relational databases

- Store information in tables in which columns are the fields and rows are the entries/ records for individual items, e.g. Excel spreadsheets, MySQL, etc.

stimulus_id	name	description	playlist	artist	album	genre	file_format	size	duration	year
1	Yeah!	iTunes Music Fragment	billboards	Usher featuring Lil' Jon & Ludacris	Confessions (Special Edition)	R&B/Soul	mp3	512606	00:00:31	2004
2	Burn	iTunes Music Fragment	billboards	Usher	Confessions (Special Edition)	R&B/Soul	mp3	512188	00:00:31	2004
3	If I Ain't Got You	iTunes Music Fragment	billboards	Alicia Keys	The Diary of Alicia Keys	R&B/Soul	mp3	512188	00:00:31	2004
4	This Love	iTunes Music Fragment	billboards	Maroon 5	Songs About Jane	Rock	mp3	512188	00:00:31	2004
5	The Way You Move	iTunes Music Fragment	billboards	OutKast & Sleepy Brown	Speakerboxxx/The Love Below	Hip-Hop/Rap	mp3	512188	00:00:31	2004
6	The Reason	iTunes Music Fragment	billboards	Hoobastank	The Reason	Alternative	mp3	511770	00:00:31	2003

Graph databases

Resource Description Framework (RDF) stores

- Don’t contain tables
  - They contain statements
- Each statement relates:
  - A **subject (s)** to an **object (o)** via a **predicate (p)**
- Statements are commonly called *triples*
- RDF triple stores are the backbone of the Semantic Web

Music-evoked remembering

- What are the basic entities that we are dealing with if we are interested in music-evoked memories?

Person, Music (Song), Memory

- Generate statements consisting of a subject (**s**), predicate (**p**), and object (**o**) that relate these three entities to each other

Example statements – in the realm of music-evoked memories

- person (s) heard (p) song (o)
- song (s) evoked (p) memory (o)
- person (s) remembered (p) memory (o)
- **Note**: We can turn these statements around by using predicates that are the **inverse of** the predicates we first used
- memory (s) evokedBy (p) song (o)
- memory (s) rememberedBy (p) person (o)
- song (s) heardBy (p) person (o)

Can we elaborate each of the previous subjects (classes) further?

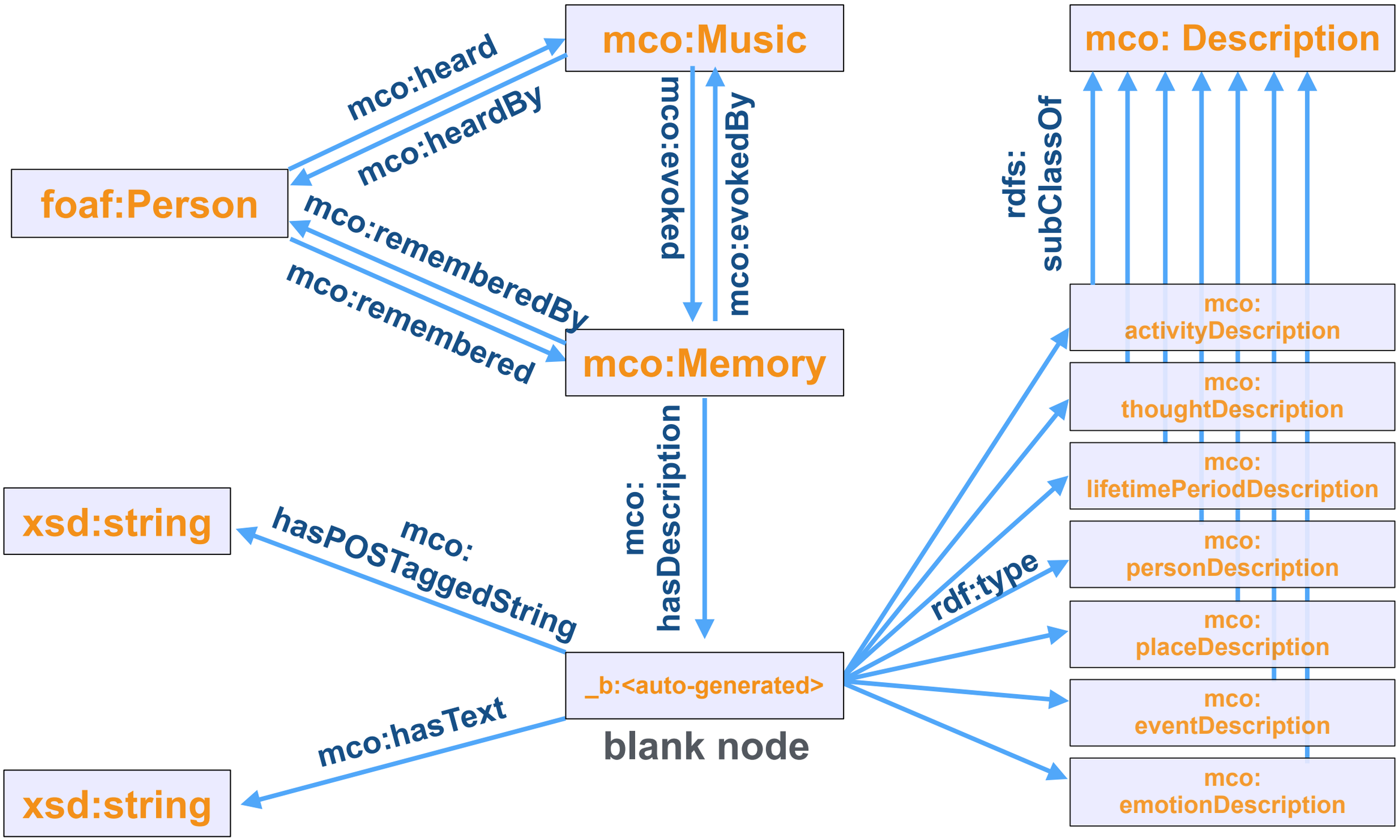
- person, song, memory

- **Exercise:**
  - For each of the entities listed above, generate a list of other entities that provides more detailed information about that entity, e.g. birthdate, title, location
  - For each of the new entities, generate a predicate that links it to the original entity (subject)
  - Write out the new statements (s, p, o)

Ontologies

- Our growing collection of statements starts to describe an **ontology** of music-evoked autobiographical knowledge
- An ontology **models** a domain of knowledge

MEAMCentral Ontology Overview



Getting data into and out of the model …

- The ontology we have discussed so far models our knowledge domain, but,
- (1) How do we get actual data into it,
  - i.e. **instances** of people hearing songs and experiencing memories?
- (2) How do we view the data we have in our database?

Semantic Web technologies

- RDF databases provide a means of specifying arbitrary models/ontologies
- Specifications exist that make it possible to store and query representations of knowledge within and across RDF databases (triple stores).
- Web Ontology Language (OWL)
  - has various extensions
- SPARQL Protocol and RDF Query Language (SPARQL)
  - a semantic web standard for querying RDF triple stores

Interfaces to Semantic Web technologies from programming, data analysis, and visualization tools

- Programming and Analysis
  - Python
  - R
- Natural Language Processing (NLP)
  - nltk (Python)
  - tm (R) - text mining
- Visualization
  - d3

Models are inherently tied to views (Part 1)

- What are the things we need to do to **get data into** our music-evoked memory **database**?
  - put information about music we want to play into the database
  - ask questions about a person
  - ask questions about a person's response for each piece of music we play them
- The questions we ask are views of the model because **each answer** we obtain **is an instance of a class** in our database. In other words, our model determines what questions we ask in order to populate our triple store with data.

Models are inherently tied to views (Part 2)

- What are the things we need to do to **view the data** stored in our music-evoked memory database?
- When we **analyze** and display data we commonly **define further relationships**/ transformations in the data in order to **answer additional questions.**
  - **Reasoning**