

Enabling Topic-based Discovery of Television Programs

Anugeetha Kunjithapatham,
Priyang Rathod, Simon Gibbs & Mithun Sheshagiri

Computer Science Lab.,
Samsung R & D, San Jose CA

Outline

- Goal
  - Background & Motivation
- Approach
  - Summary
  - Details
    a. Enriching Video Metadata
    b. Building Topic Taxonomy
    c. Categorizing Videos
- Prototype: ‘DeepVideo’
  - UI Screenshots
- Conclusion
  - Future Work
  - Related Work

Goal

- Enable Intuitive Topic-based Discovery of Videos
  - Users choose a topic of interest from a hierarchy of topics
    - Browse videos related to topic
    - or… Choose a sub-topic of topic
      - Browse videos related to the sub-topic
      - or… Choose a sub-topic of the sub-topic
        - » and so on…

- Example Scenario:
  - "User wants to explore US politics and learn about famous presidents"

Politics -> Politicians -> Presidents

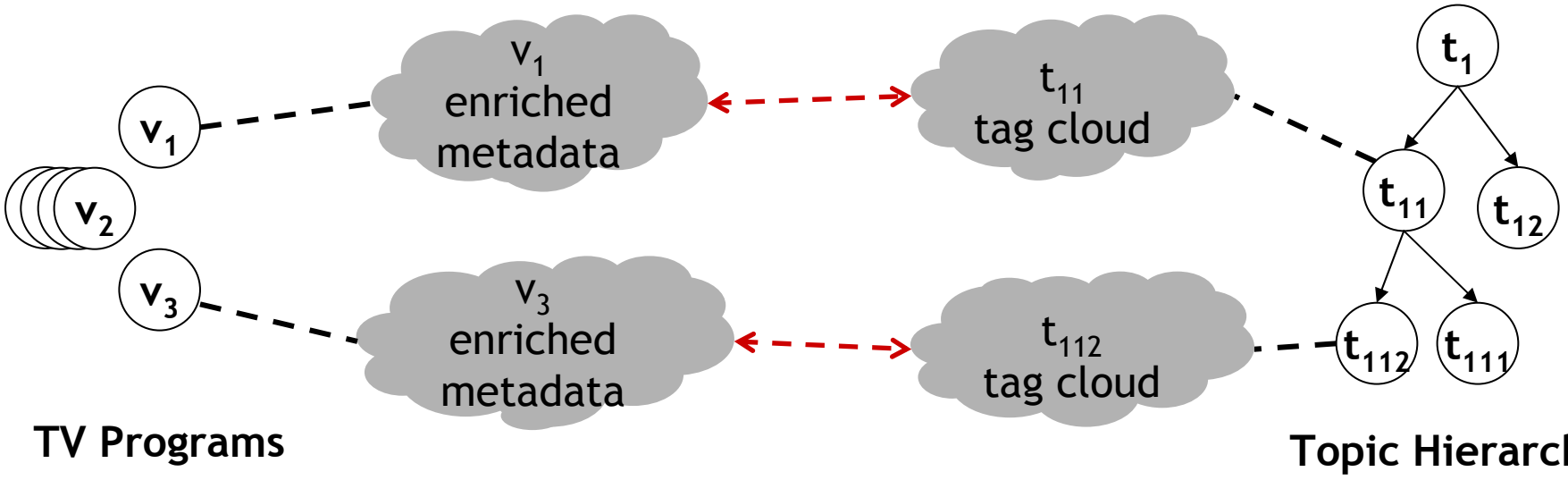
-> Documentaries about ‘John F. Kennedy’

Background & Motivation

- TV Programs continue to be popular
  - 100-fold increase in programs broadcast in the last 50 yrs (TMS)
  - Huge archive of Documentary videos, with immense educational value
- Discovering programs of interest not easy
  - Limitations with EPG metadata-based search
    - Ambiguous, misleading Genres
      - Genre '*Travel*' *for* "Earth Diaries: Hindu prayer offerings at Ganges…"
      - '*History*', '*Documentary*' are high-level
    - Vague Keywords
      - No indication of type (Person, Location)
      - No info on related attributes
    - Keyword-based search not efficient
      - Does not scale well (TiVo)
      - Different words refer to same entity / related concepts
        - » "Traditional pasta from Tuscany" doesn't turn up for 'Italian Cuisine'

Approach

- Semantically enrich program metadata
- Build a taxonomy of topics and concepts
  - Personalize: rank keywords/topics related to watched / favorite videos higher
- Map / Categorize Videos to Topics



a. Enriching Video metadata

1. Significant keywords/phrases identified from Metadata (Genre-based NLP; Messer, 2008), ranked based on:
   - keyword frequency
   - type & # of fields (title, desc, closed-captions) it appears in
2. Keywords mapped to known 'Topics'
   - Freebase, TMS archives used as topic sources
   - String distance to identify/fix possible typos
3. Type & attributes of Topic(s) identified
   - Considered secondary topics for the video

➢ Say, a program's Description is: **"Traditional Pasta from Tuscany".**
   - Topics associated would include: ***Italy***, ***Tuscany***, ***Pasta***, ***Cuisine***, ***Italian***

b. Building Topic Taxonomy

- Seed taxonomy consisting of topics (‘nodes’), sub-topics (‘child nodes’) & keywords
  - Structure/ subject known; Keywords to be determined
  - Node: <name, children, keywords>
- Identify keywords (or, concepts) related to each node
  - Use multiple Internet sources
    - Freebase & Wikipedia; Google, Yahoo & Mozilla Directories; WordNet & Dictionary.com; Ask, Google & Yahoo Search Engines
  - Keywords ranked based on:
    - # of sources it appears in, # of times it appeared in each source, reliability of the source
    - if it is related to a video watched by the user in the past (personalization)

c. Categorizing Videos

- Videos categorized by topics in Taxonomy
  - Video Keywords are matched with Topic Keywords
  - Degree of relevance of topic to video (Video-Topic Rank) is determined based on:
    - # of common keywords
    - rank of topic keyword & video keyword

$$vtRank_0(V,T) = \sum_{i=1}^{n} vkRank(V, K_i(T))$$

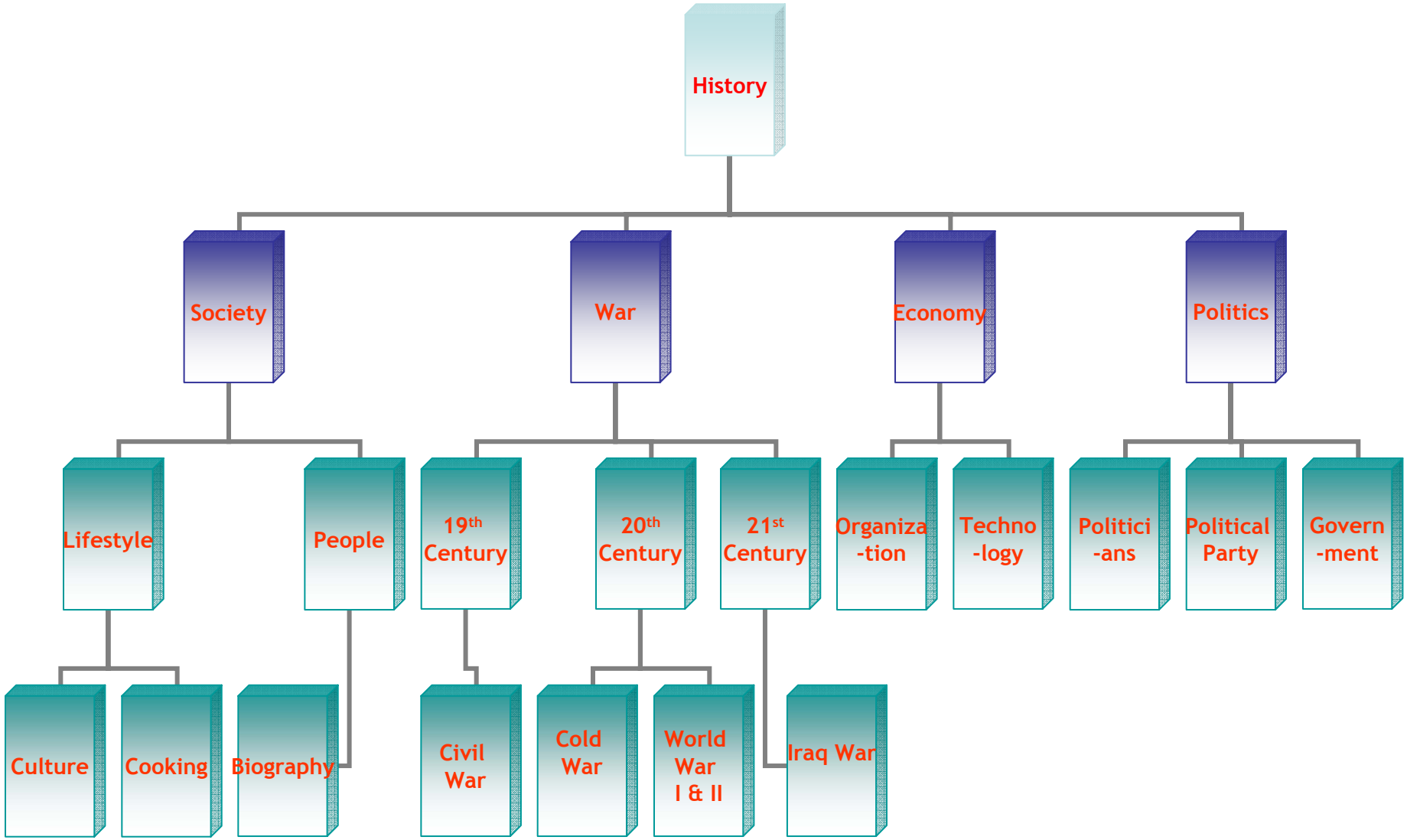
    - if video is related to parent-topics of related topics & the hierarchical distance between topic/parent-topic

$$vtRank(V,T) = \sum_{i=1}^{n} \frac{vtRank(V, PT_i(T))}{D(T, PT_i(T))} + vtRank_0(V,T)$$

Prototype

- Modules:
  - Taxonomy builder, metadata enricher, video categorizer and the UI
- Tools:
  - .NET technologies
    - UI module implemented in C# and XAML
    - Other 3 modules implemented in C#
  - SQL server for data storage and processing
- Data:
  - 30K Documentary TV programs
    - Genre: ‘History’ & ‘Documentary’ (Historical Documentaries)
  - Metadata obtained from TMS EPG Archives
    - Title, description, genres, sub-title & episode name
- Seed taxonomy with ‘History’ as the root topic
  - Enriched with keywords (described in slide 5)

Taxonomy


History
Society
War
Economy
Politics
Lifestyle
People
19th Century
20th Century
21st Century
Organiza -tion
Techno -logy
Politici -ans
Political Party
Govern -ment
Culture
Cooking
Biography
Civil War
Cold War
World War I & II
Iraq War

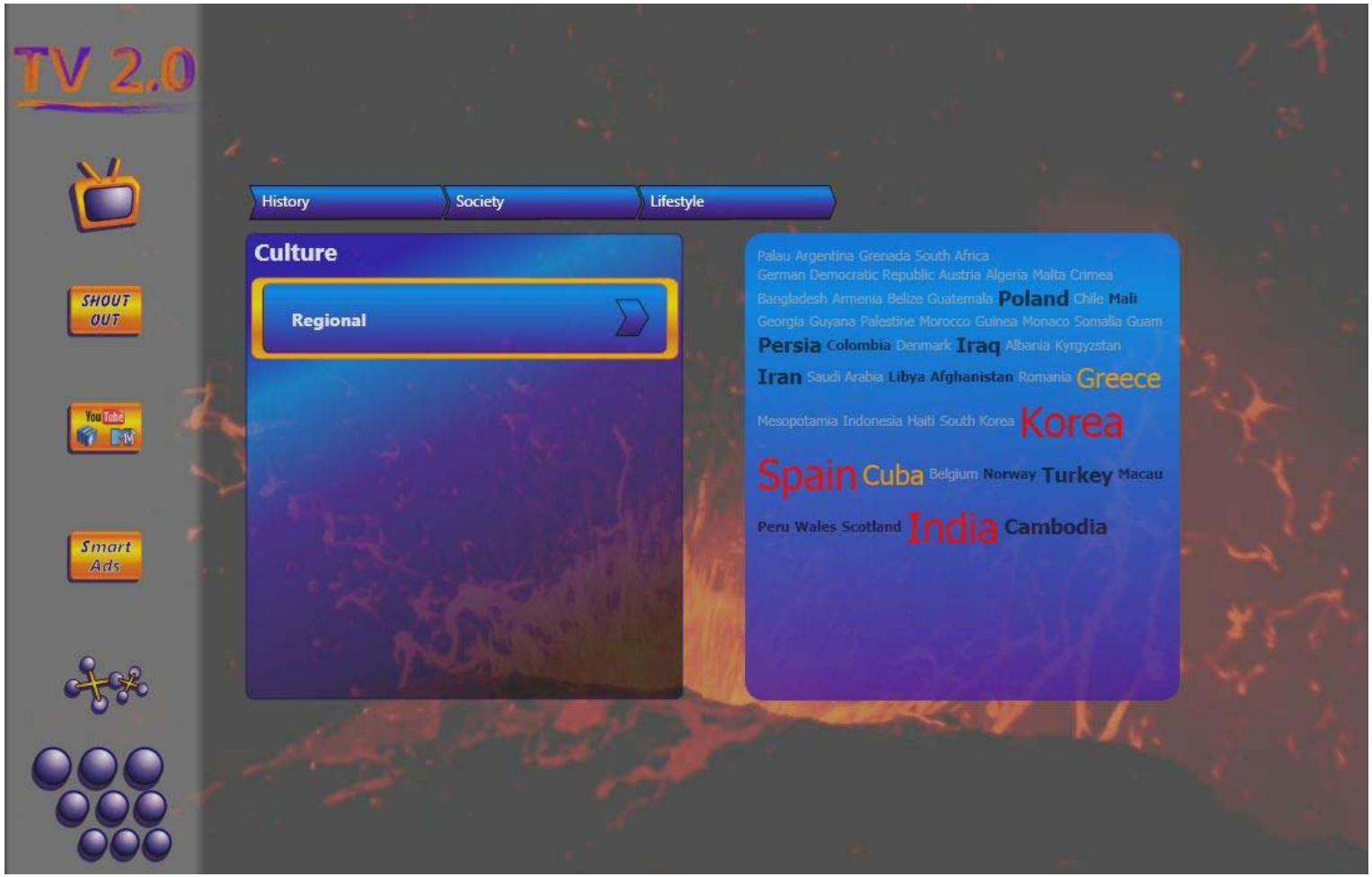
UI Screenshots - 1



UI Screenshots - 2



UI Screenshots - 4



Conclusion

- Explored initial ideas:
  - Enable intuitive discovery of videos based on topic of interest
  - Built quick & practical prototype to:
    - enrich metadata, create topic taxonomy, categorize videos
- Possible applications:
  - Value-addition to the state-of-the-art STBs/TVs
    - Enhance the naïve EPG-based navigation
    - Better keyword-based search
  - New kind of interface to explore a structured collection of content

Future Work

- Seed Taxonomy built manually
  - Automate taxonomy creation
  - Consider taxonomies with multiple hierarchies
- Metadata Enrichment: 1st keyword-topic match is assumed
  - Implement a topic verification step
  - Evaluate overlapping attributes for all topics corresponding to metadata keywords
- Positive feedback from internal Tech. Fair
  - Formal user evaluation to be done

Related Work Discussion

- Prior Work:
  - YouTube, Flickr, Del.ici.ous: user chosen tags, category
  - Twine, Faviki: use Wikipedia to infer relationships b/w content
  - Smoliar'94, Yuan'06, Wang'07,Yang'07 & others: categorize videos

  Summary:
  - User-initiated or semi-automatic categorization of content through clustering & classification techniques
  - Simply based on words appearing in the content metadata or text

- **Differences to our Approach**
  - Metadata enrichment
    - Freebase: vast, structured & semantic resource for topics
  - Degree of relevance between video & topic is critical
    - Ensures most relevant videos are ranked higher for a chosen topic
  - Personalization of video-topic relevance
    - no user intervention

Thank You!