



A Large Scale Concept Ontology for Multimedia Understanding

**Milind Naphade, John R. Smith, Alexander Hauptmann, Shih-Fu Chang &
Edward Chang**

IBM Research, Carnegie Mellon University, Columbia University & University of California at Santa Barbara
naphade@us.ibm.com jsmith@us.ibm.com alex@cs.cmu.edu sfchang@ee.columbia.edu
echang@xanadu.ece.ucsb.edu



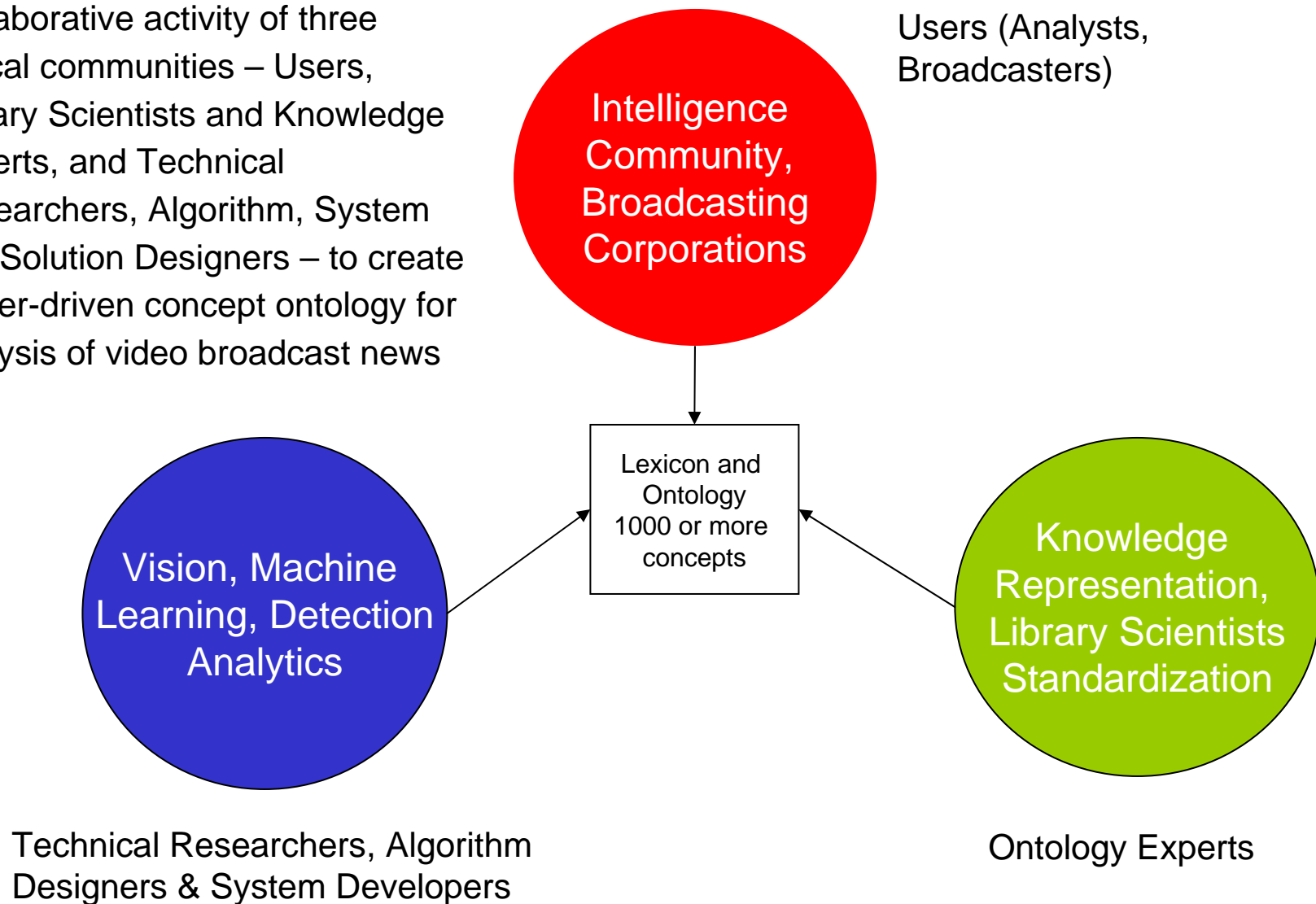
NRRC
MITRE

April 2005

NWRRC
Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

Central Idea

- Collaborative activity of three critical communities – Users, Library Scientists and Knowledge Experts, and Technical Researchers, Algorithm, System and Solution Designers – to create a user-driven concept ontology for analysis of video broadcast news



Problem

- **Users and analysts require richly annotated video content for accomplishing required access and analysis functions over massive amount of video content.**
- **Big Barriers:**
 - **Research community needs to advance technology for bridging gap from low-level features to semantics**
 - **Lack of large scale useful well-defined semantic lexicon**
 - **Lack of user-centric ontology**
 - **Lack of corpora annotated with rich lexicon**
 - **Lack of feasibility studies for any ontology if defined**
- **Examples:**
 - **The TRECVID lexicon defined from a frequentist perspective. Its not user-centric.**
- **No effort to date to design lexicon by joint partnership between different communities (users, knowledge experts, technical)**

Workshop Goals

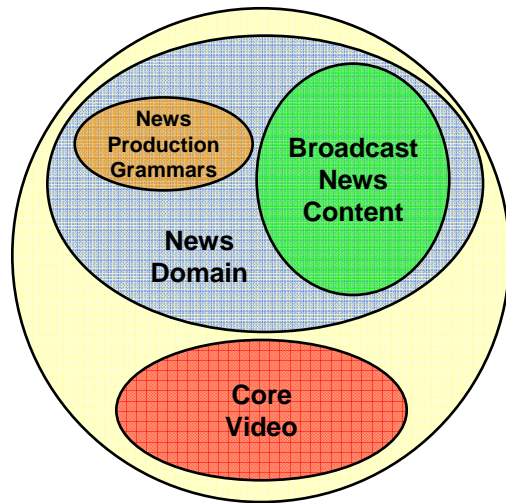
- **Organize series of workshops that bring together three critical communities – Users, Library Scientists and Knowledge Experts, and Technical Researchers – to create a ontology on order of 1000 concepts for analysis of video broadcast news**
- **Ensure impact through focused collaboration of these different communities to achieve balance of usefulness, feasibility and size**
- **Specific Tasks:**
 - **Solicit input on user needs and existing practices**
 - **Analyze applications, prior work, concept modeling requirements**
 - **Develop draft concept ontology for video broadcast news domain**
 - **Solicit input on technical capabilities**
 - **Analyze technical capabilities for concept modeling and detection**
 - **Form benchmark and define annotation tasks**
 - **Annotate benchmark dataset**
 - **Perform benchmark concept modeling, detection and evaluation**
 - **Analyze concept detection performance and revise concept ontology**
 - **Conduct gap analysis and identify outstanding research challenges**

Workshop Format and Duration

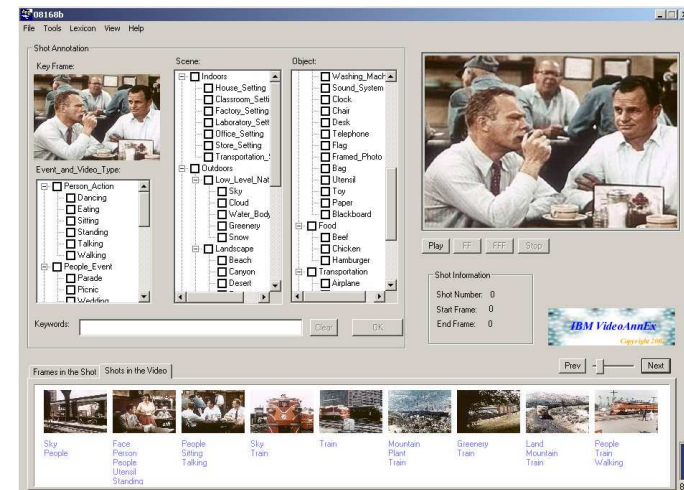
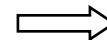
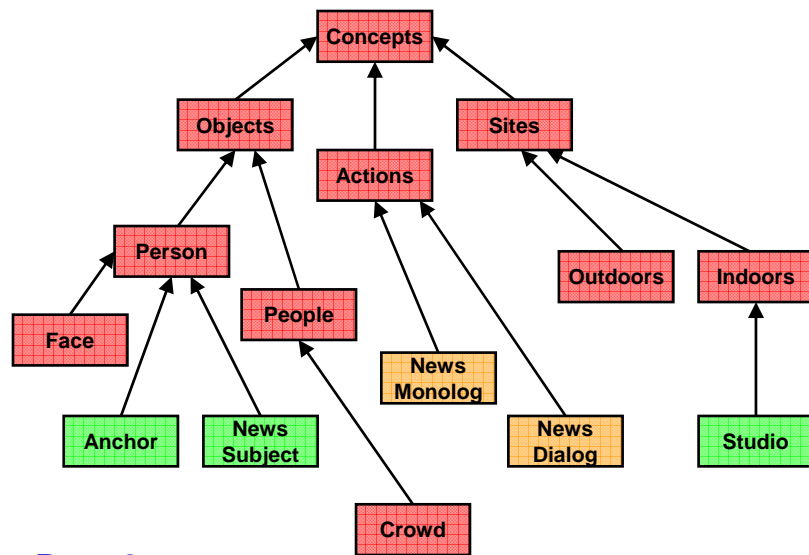
- **Propose to hold two multi-week workshops accompanied by annotation, experimentation, and prototyping tasks**
- **Focus on video broadcast news domain**
- **Workshop Organization:**
 - **Pre-workshop 1: *Call for Input on User Needs and Existing Practices***
 - **Ontology Definition Workshop (two-weeks):**
 - **Part 1: *User Needs***
 - **Part 2: *Technical Analysis***
 - **Ad hoc Tasks**
 - **Task 1: *Annotation***
 - **Task 2: *Experimentation***
 - **Task 3: *Evaluation***
 - **Ontology Evaluation Workshop (two-weeks):**
 - **Part 1: *Validation and Refinement***
 - **Part 2: *Outstanding Challenges and Recommendations***
- **Substantial off-line tasks for annotation and experimentation require organization as two separate workshops**

Broadcast News Video Content Description Ontology

Broadcast News Ontology



- Why the Focus on Broadcast News Domain?
 - Critical mass of users, content providers, applications
 - Good content availability (TRECVID, LDC, FBIS)
 - Shares large set of core concepts with other domains
- Ontology Formalism:
 - Entity-Relationship (E-R) Graphs
 - RDF, DAML / DAML+OIL, W3C OWL
 - MPEG-7, MediaNet, VEML
- Seed Representations:
 - TRECVID-2003 News Lexicon (Annotation Forum)
 - Library of Congress TGM-I
 - CNN, BBC Classification Systems



MPEG-7 Video Annotation Tool

Approach (Pre-workshop and 1st workshop)

- **Pre-workshop: Call for Input**
 - Solicit input on user needs and existing practices
- **Ontology Definition Workshop**
 - Part 1: *User Needs***
 - Analyze use cases, concept modeling requirements, prior lexicon and ontology work
 - Develop draft concept ontology for video broadcast news domain
 - Output: Version 1**
 - Requirements and Existing Practices
 - Domain Concepts and Ontology System
 - Video Concept Ontology
 - Part 2: *Technical Analysis***
 - Analyze technical capabilities for concept modeling and detection
 - Form benchmark and define annotation tasks
 - Output: Version 1**
 - Benchmark (Use cases, Annotation)

Approach (Ad-hoc Tasks and 2nd workshop)

- **Ad hoc Group**

- Task 1: *Annotation*

- Annotate benchmark dataset

- Task 2: *Experimentation*

- Perform benchmark concept modeling and detection

- Task 3: *Evaluation*

- Evaluation of concept detection, ontology and use of automatic detection for use cases and evaluation

- Output:**

- Benchmark v.2
 - Concept Detection Evaluation v.1
 - Ontology Evaluation v.1
 - Query Answering Effectiveness with Automated Detection Evaluation v.1

- **Ontology Evaluation Workshop**

- Part 1: *Validation*

- Analyze evaluation of ontology, concept detection and its application to use case answering.

- Output**

- Domain Concepts v.2 and Ontology System v.2
 - Video Concept Ontology v.2

- Part 2: *Outstanding Challenges*

- Conduct gap analysis and identify outstanding research challenges

- Output:**

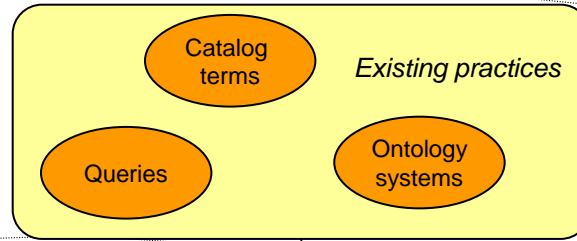
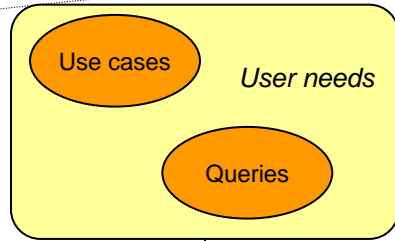
- Research Challenges v.1

Users:

- IC analysts
- Broadcasters

Ex. Usage:

- Searching
- Browsing
- Navigation
- Threading
- Summarization



Ex. Ontology & catalog systems:

- RDF
- Topic Maps
- DAML / DAML+OIL
- W3C OWL
- Dublin Core
- OCLC
- Open Directory Project
- MARC
- MODS
- MediaNet
- VEML

Ex. Catalog terms

- LOC TGM-I
- MPEG-7 classif. Schemes

Ex. Queries

- TRECVID
- BBC

**Pre-Workshop:
Call for Input**

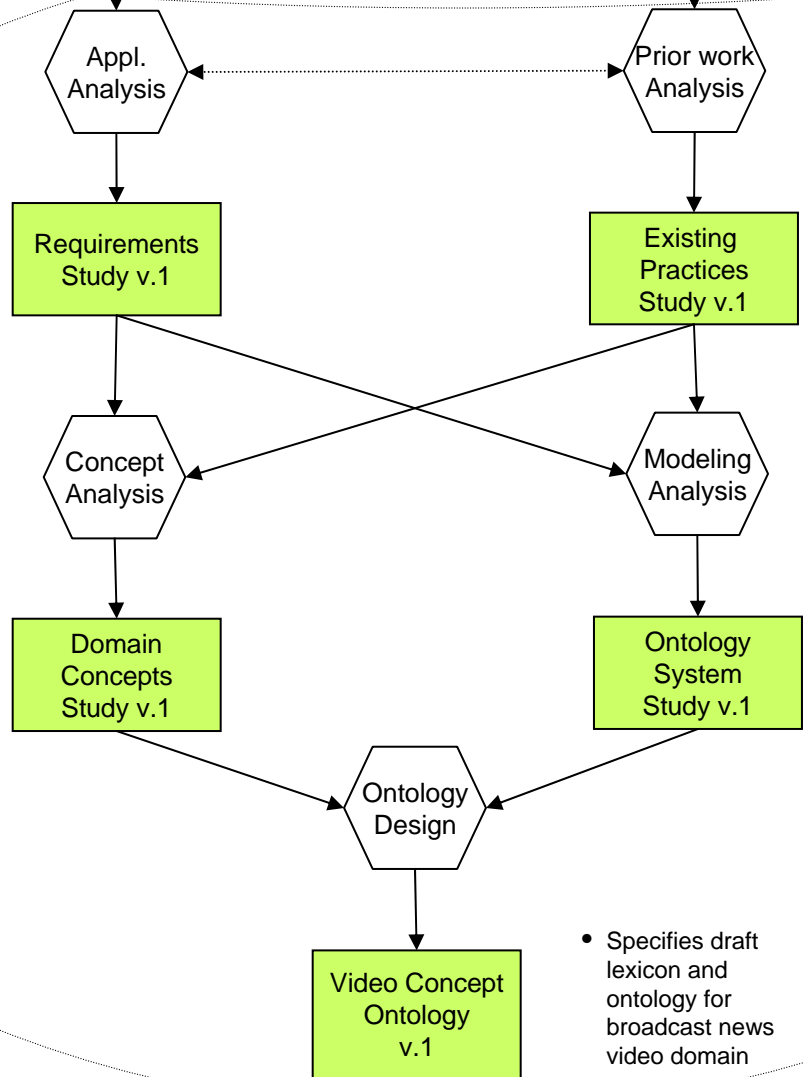
- Identifies indexing requirements for broadcast news video including example queries

- Identifies and defines domain concepts, terms, classification systems for broadcast news video (ex. objects, actions, sites, events)

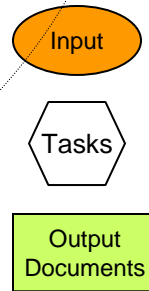
- Documents existing practices for indexing broadcast news

- Specifies ontology system for broadcast news video domain

**Workshop 1:
User Needs**

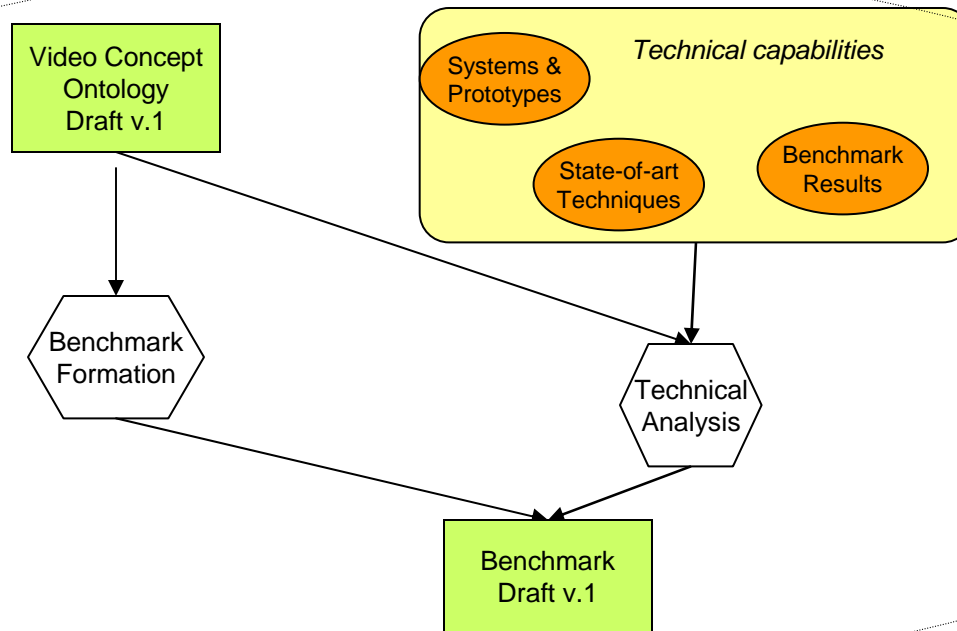


- Specifies draft lexicon and ontology for broadcast news video domain



Workshop 1: Technical Analysis

- Defines benchmark for concept detection for video broadcast news (dataset, detection and search tasks, metrics)



Systems:

- Video logging
- Video retrieval

Ex. State-of-art techniques:

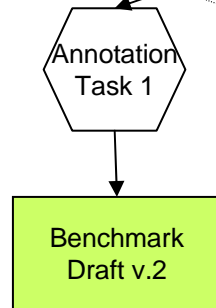
- Content-based search
- Segmentation
- Tracking
- High-level feature detection
- Story segmentation

Ex. Benchmarks:

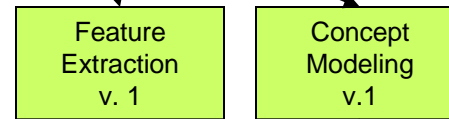
- TRECVID

Ad Hoc Task 1: Annotation

- Refines benchmark to include annotated ground-truth for experimentation and evaluation

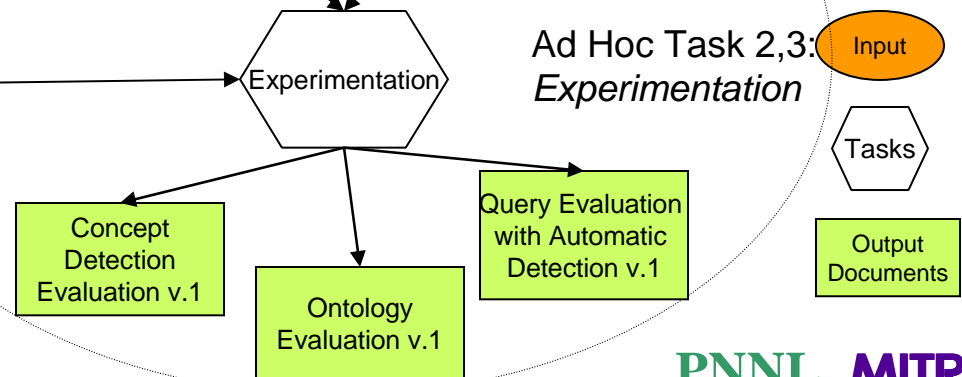


- Identifies features required for modeling each of draft concepts



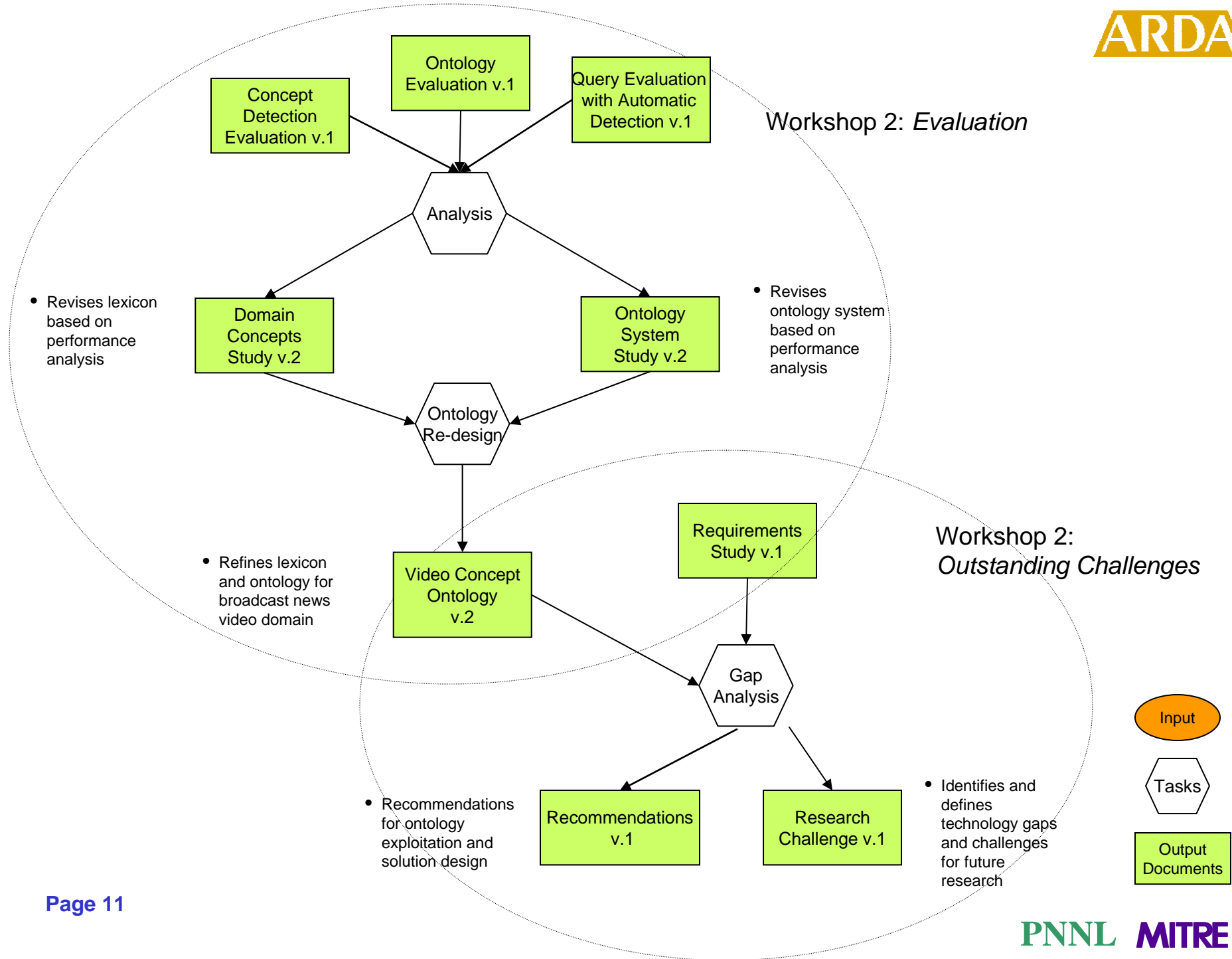
- Identifies and maps techniques for modeling and detecting each of draft concepts

Ad Hoc Task 2,3: Experimentation



Workshop 2: Evaluation

Workshop 2: Outstanding Challenges



Domain and Data Sets

- **Candidate data set:**

- TRECVID Corpus (>200 hours of video broadcast news from CNN and ABC). Has the following advantages
 - availability
 - generalization capability better with than other domains
 - # of research groups up to speed on this domain for tools/detectors
 - TREC established some benchmark and evaluation metrics already.
- Will avoid letting domain specifics influence the design of ontology to an extent where the ontology starts catering to artifacts of the BN domain.
- Will seek other sources such as FBIS, WNC etc.

- **Annotation issues:**

- Plan to leverage prior video annotation efforts where possible (e.g., TRECVID annotation forum)
- Hands-on annotation effort will induce discussions and requires refinements of concepts meanings

Evaluation Methods

- **Require benchmarks and metrics for evaluating:**
 - **Utility of ontology – coverage of queries in terms of quality and quantity**
 - **Feasibility of ontology:**
 - **Accuracy of concept detection and degree of automation (amount of training)**
 - **Effectiveness of query systems using automatically extracted concepts**
- **Metrics of Retrieval Effectiveness**
 - **Precision & Recall Curves, Average Precision, Precision at Fixed Depth**
- **Metrics of Lexicon Effectiveness**
 - **Number of Use Cases that can be answered by lexicon successfully**
 - **Mean average precision across the set of use cases**
- **Evaluate at multiple levels of granularity:**
 - **Individual concept, classes, hierarchies**

Confirmed Participants – Knowledge Experts and Users

Library Sciences and Knowledge representation (definition of lexicon):

- Corrine Jorgensen, School of Information Studies, Florida State University
- Barbara Tillett, Chief of Cataloging Policy and Support, Library of Congress
- Jerry Hobbs, USC / ISI
- Michael Witbrock, Cycorp
- Ronald Murray, Preservation Reformatting Division, Library of Congress

R&D Agencies

- John Prange, ARDA
- Sankar Basu, Div. of Computing and Comm. Foundations, NSF
- Maria Zemankova, Div. of Inform. and Intell. Systems., NSF

Standardization and Benchmarking (theoretical and empirical evaluation):

- Paul Over, NIST
- John Garofolo, NIST
- Donna Harman, NIST
- David Day, MITRE
- John R. Smith, IBM Research

User Communities (interpretation of use cases for lexicon definition, broadcasters help getting query logs for finding useful lexical entries)

- Joanne Evans, British Broadcasting Corporation
- Chris Porter, Getty Images
- ARDA and analysts

Confirmed Participants – Technical Team

Theoretical Analysis:

(Help conduct analysis during initial lexicon and ontology design)

- Milind R. Naphade, IBM Research
- Ramesh Jain, Georgia Institute of Technology
- Thomas Huang, UIUC
- Edward Delp, Purdue University

Experimentation: (Help address evaluation issues for lexicon, ontology and concept evaluation)

- Alexander Hauptmann, CMU
- Alan Smeaton, Dublin City University
- HongJiang Zhang, Microsoft Research
- Ajay Divakaran, MERL
- Wessel Kraaij, Information Systems Division, TNO TPD
- Ching-Yung Lin, IBM Research
- Mubarak Shah, University of Central Florida

Prototyping: (Help with prototyping tools for annotation, evaluation, querying, summarization and statistics gathering)

- Shih-Fu Chang, Columbia University
- Edward Chang, UCSB
- Nevenka Dimitrova, Phillips Research
- Rainer Lienhart, Intel
- Apostol Natsev, IBM Research
- Tat-Seng Chua, NUS
- Ram Nevatia, USC
- John Kender, Columbia University

Impact and Outcome

- First of a Kind Ontology of 1000 or more semantic concepts that have been evaluated for their usability and feasibility by different communities including UC, OC, MC.
- Annotated corpus (200 hours) and ontology can be further exploited for future TRECVID, VACE, MPEG-7 activities. Core semantic primitives, that can be included in various video description standards/languages such as MPEG-7.
- Empirical and theoretical study of automatic concept detection performance for elements of this large ontology. Use of current state of the art detection wherever possible. Use of simulation where the detection is not available.
- Use cases (queries) testing and expansion into ontology
- Reports documenting use cases, existing practices, research challenges and recommendations
- Prototype systems and tools for annotation, query formulation and evaluation
- Guidelines on manual and automatic multimedia query formulation techniques going from use-cases to concepts.
- Categorization of classes of concepts based on feasibility, detection performance and difficulty in automation

BOTTOMLINE: All this is driven by the user

Summary of Key Questions

- **How easy was it to create annotations**
 - (man-hours/hr of video?)
- **How well does the lexicon 'partition' the collection**
- **Given perfect annotations/classification:**
 - How well does the lexicon aid with queries/tasks
- **How good is automatic annotation of the sample collection**
 - What fraction of perfect annotations accuracy is obtained for the queries/tasks
- **How much is automatic classification performance of a given lexical item a function of training data**
 - Estimate how much training data would get this lexical item to 60%, 80%, 90%, 95%?
- **What lexicon changes are necessary or desirable?**
- **Are 1000 concepts the right ballpark?**
- **What are shortcomings of ontology-driven approach?**

Video Event Ontology (VEO) & VEML

- A Video Event Ontology was developed in the *ARDA workshop on video event ontologies for surveillance and meetings* allows natural, hierarchical representation of complex spatio-temporal events common in the physical world by a composition of simpler (primitive) events
- VEML – XML-derived Video Event Markup Language used to annotate data by instantiating a class defined in that ontology. Example: We will attempt to use or adapt their notation to the extent possible
- (<http://www.veml.org:8668//space/2003-10-08/StealingByBlocking.veml>)
- Broadcast video news ontology is likely to have little overlap with the complex surveillance events described in the VEO, except for some basic concepts. We expect our ontology to be broader, but much shallower
- Our broadcast news ontology is *largely* applicable to any edited broadcast video (e.g. documentaries, talk shows, movies) and *somewhat* applicable to video in general (including surveillance, UAV and home videos).