FXPAL

FXPAL Interactive Search for TRECVID 2004

John Adcock, Matthew Cooper,
Andreas Girgensohn, Lynn Wilcox

Overview

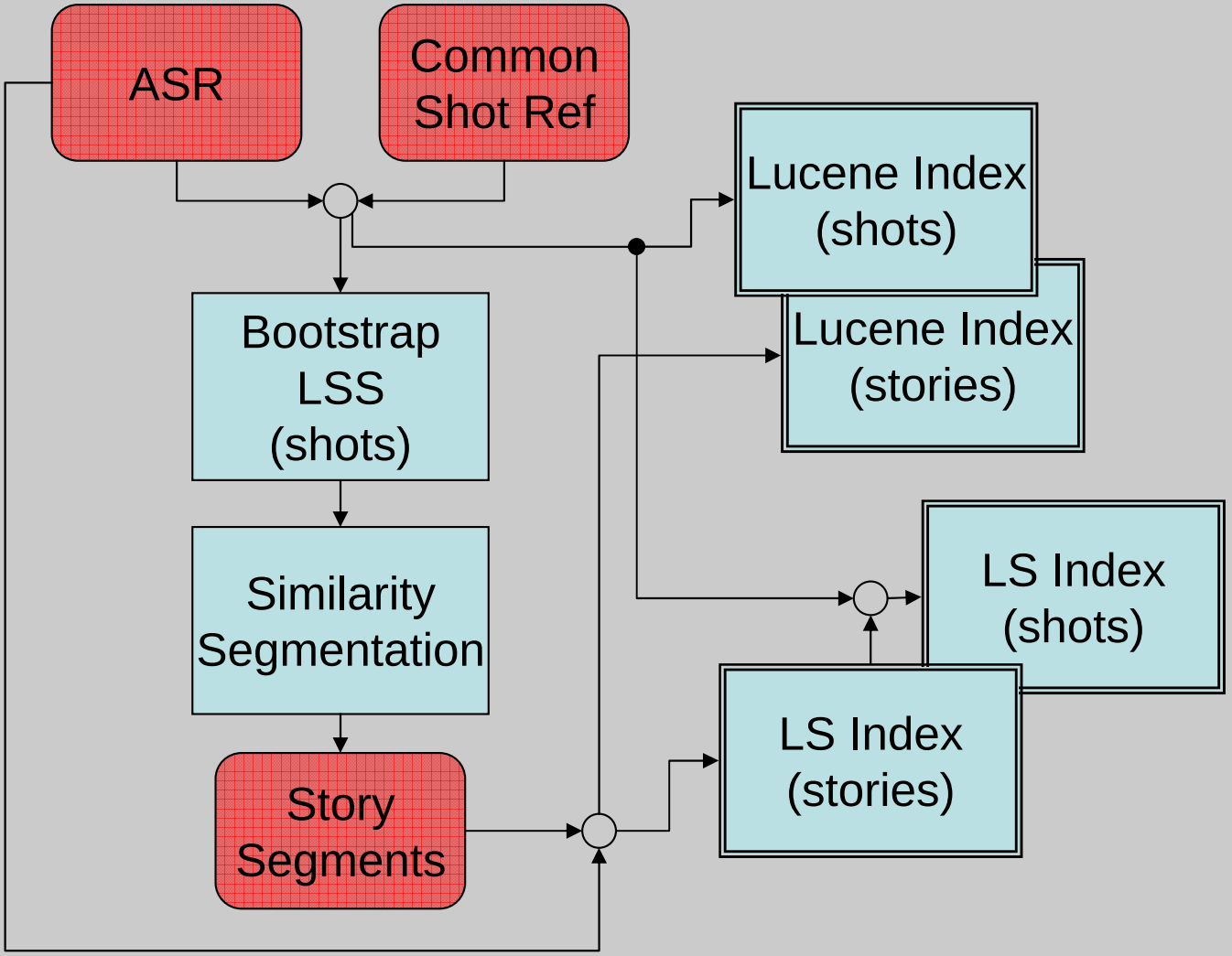
- First time doing search
  - 2nd year of participation overall
- Emphasis on interface elements
  - Rich visualization of search results
  - Quick and easy exploration of results
- Straightforward search engine
  - Text search over ASR transcripts
    - Literal search with Lucene
    - Fuzzy search with LSS
  - Keyframe search by image similarity
    - Color correlograms

Preprocessing

Unit of search retrieval is a “story”, but we couldn’t don’t have reference story segmentation for the test set

- Group reference shots into “stories”
  - Bootstrap an LSS with common shot boundaries and ASR
  - use similarity-matrix method to find “story” boundaries
- Given new story boundaries
  - Generate text indices for story and shots
  - Generate story-based LSS for search

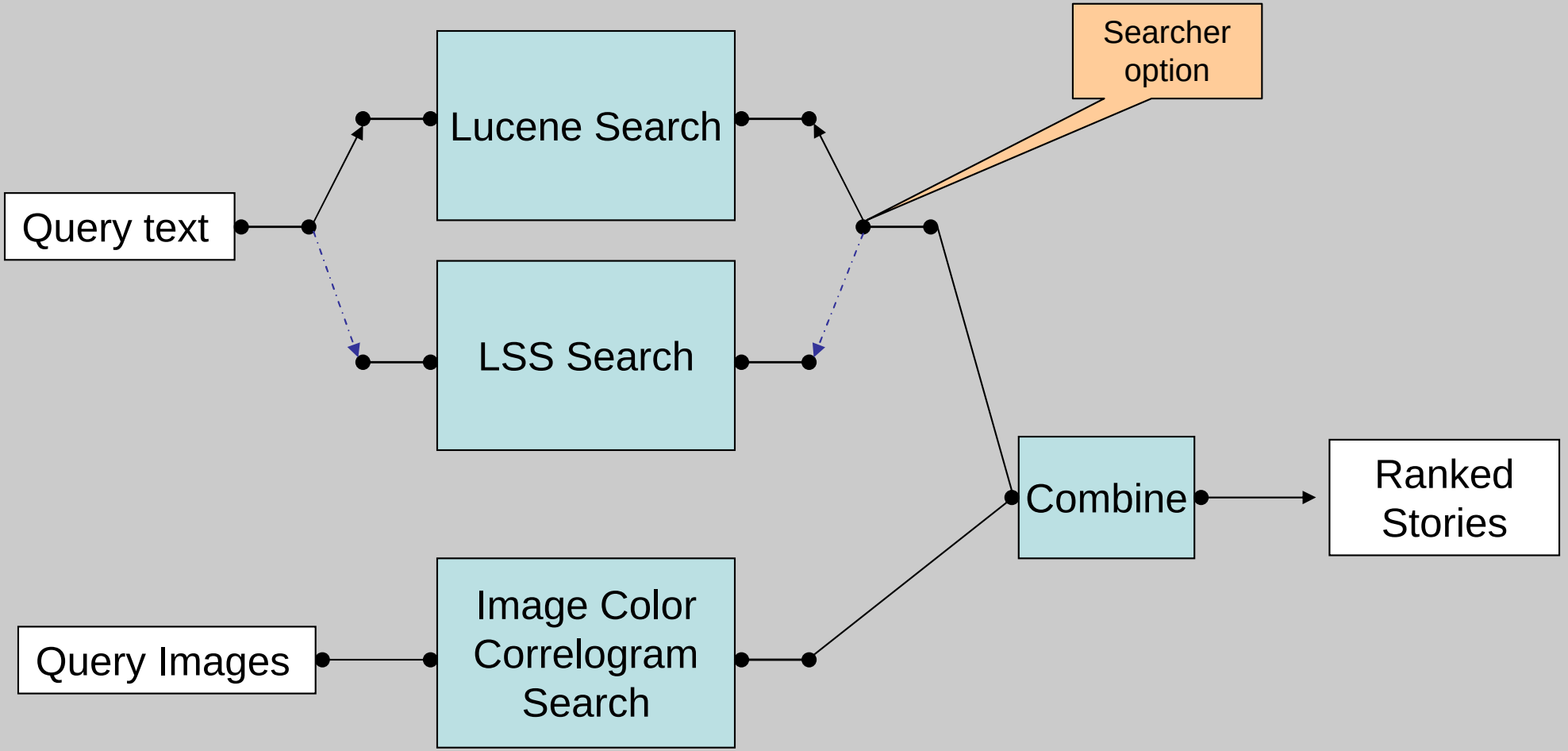
Preprocessing



Search Engine

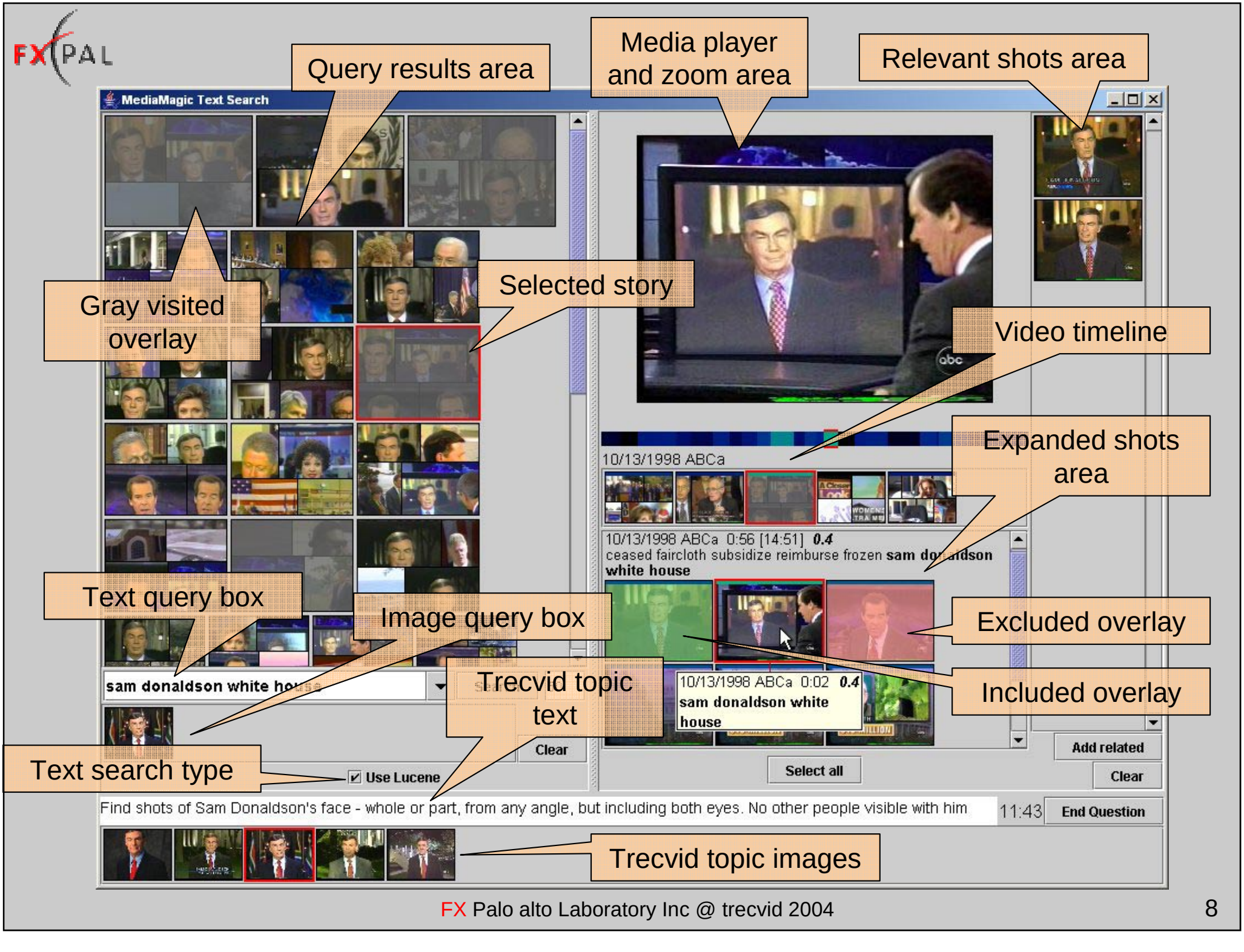
- User specifies combination of:
  - Text query
    - Literal query using Lucene *or* fuzzy query using LSS
  - Image examples
    - Any keyframe in the interface can be dragged onto the image example area
  - Text/image weighting is static and equal
  - Max image similarity of shot propagated to story
  - Text similarity of story propagated to shot
    - Averaged with shot-based text similarity

Search Engine



Interface Elements

- Stories summarized in keyframe “quads”
- Navigate through stories to video timeline/shots
- Transparent icon overlays
  - Visited: grayed
  - Relevant: green
  - Irrelevant:red
- Query-relevance shown with size and color
- Hotkeys for most actions
- Multi-select and drag and drop


FX PAL
Query results area
Media player and zoom area
Relevant shots area
MediaMagic Text Search
Gray visited overlay
Selected story
Video timeline
10/13/1998 ABCa
Expanded shots area
10/13/1998 ABCa 0:56 [14:51] 0.4
ceased faircloth subsidize reimburse frozen sam donaldson white house
Text query box
Image query box
Excluded overlay
sam donaldson white house
10/13/1998 ABCa 0:02 0.4
sam donaldson white house
Trecvid topic text
Included overlay
Clear
Add related
Select all
Clear
Use Lucene
Text search type
Find shots of Sam Donaldson's face - whole or part, from any angle, but including both eyes. No other people visible with him
11:43
End Question
Trecvid topic images

Story Summary Quads

- Query-dependent story summary
  - Use 4 highest scoring shots in the story
  - Allocate space proportional to score

Story thumbnail

Shot thumbnails

Building on searches

- Find similar
  - Use shot/story text for search
- Add related
  - Auto re-query with existing results

Expanded Story / Timeline Browsing

- Selecting a story expands the video at that point
  - Clickable video timeline with relevancy shading
  - Clickable story quad timeline
  - Shot thumbs marked with relevancy
  - Overlay on shots marked (non)relevant
  - Mouse-overs zoom in the media player and tool-tip shows relevancy context
  - Double clicks play video in the media player

Experiments

- 6 searchers answering 12 topics each in latin square
  - Pairs of orthogonal users grouped together
    - Each topic answered 3 times
  - Searchers include 2 primary developers
    - 1 ended up in best and 1 in worst performing group
- Each of the 3 complete searcher runs goes through 3 “systems” or methods for filling out the shot list yielding 9 total submissions

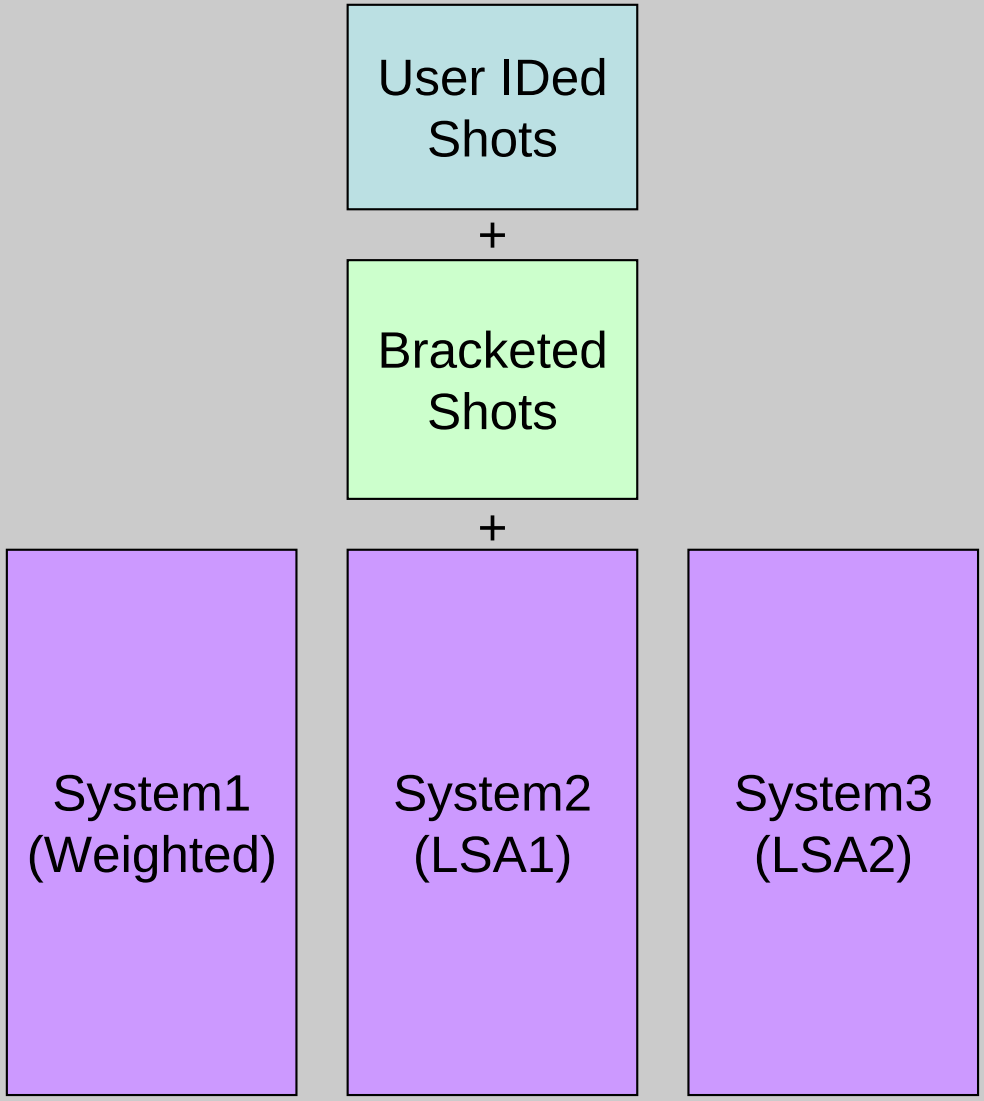
System Types

- Type 1:
  re-issue user queries and weight results of each query by precision against the user-labeled shots
- Type 2:
  take text from all relevant shots and issue a single new LSS-based text query
- Type 3:
  take text from each relevant shot in turn for LSS-based query and apply query ranking as in system type 1

Shots marked as not-relevant excluded from system results

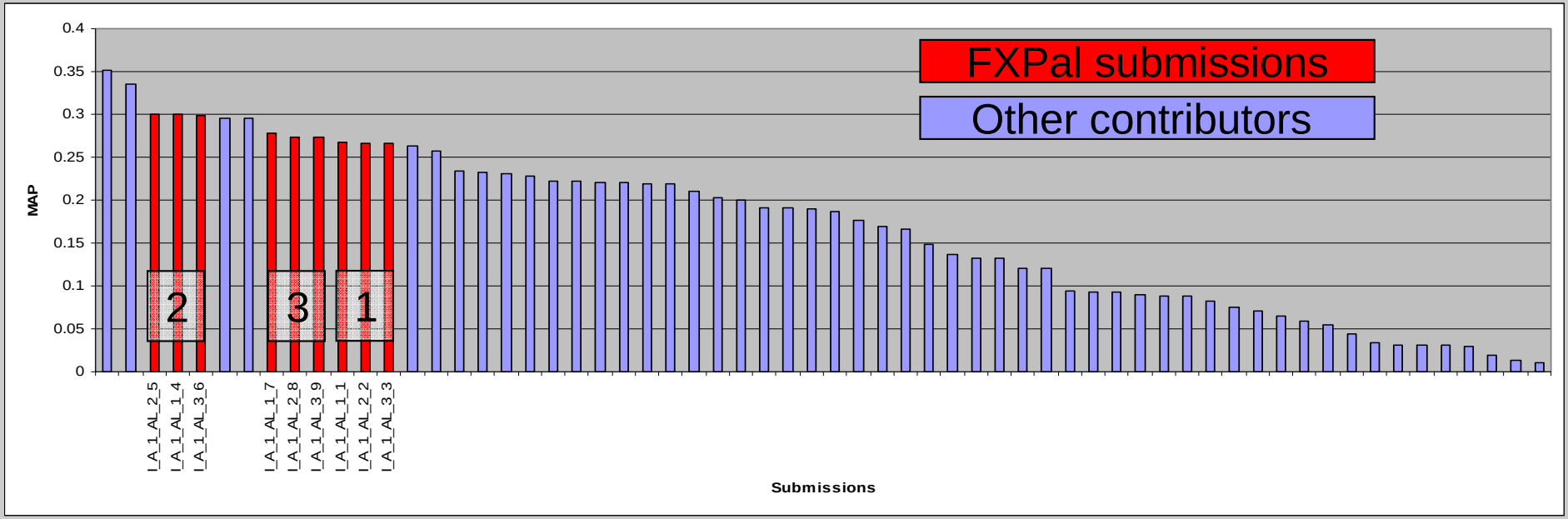
Every system type preceded by bracketing the user-retrieved shots

Submissions

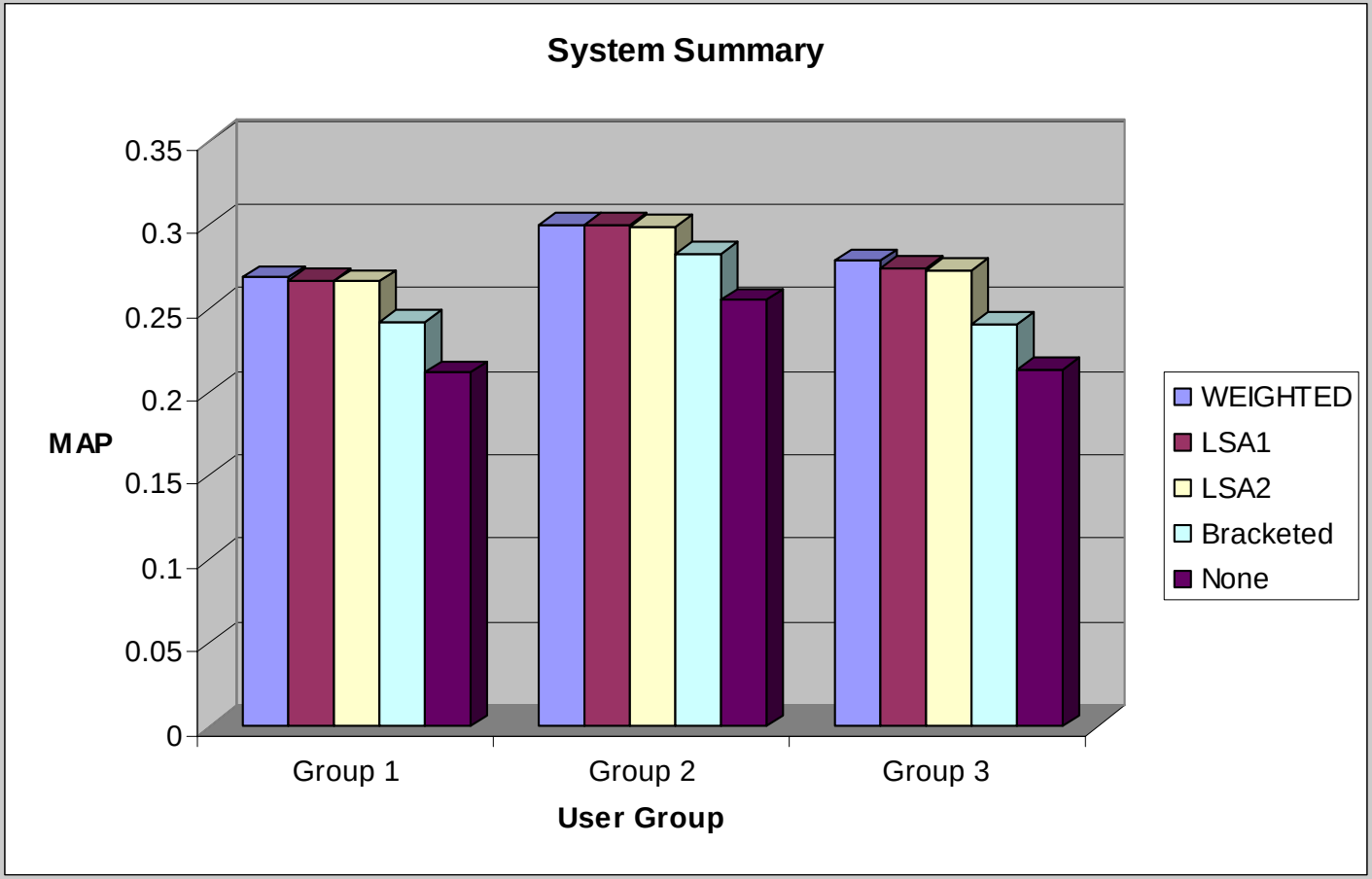


Results

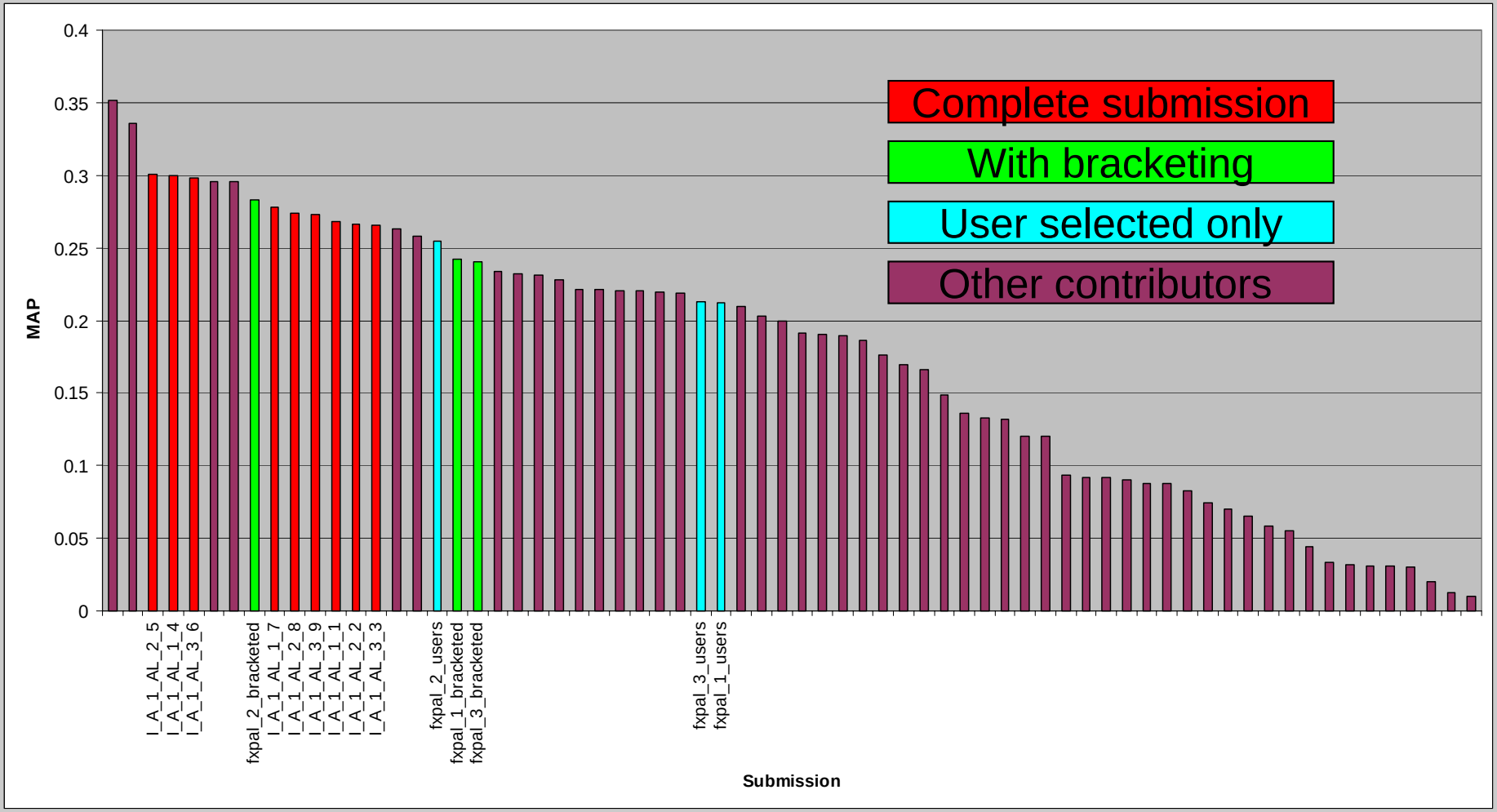
- Ranks 3-6, 9-13 in overall MAP
  - Strongly user dependent (user groups clump together)
  - Post-processing methods perform nearly same



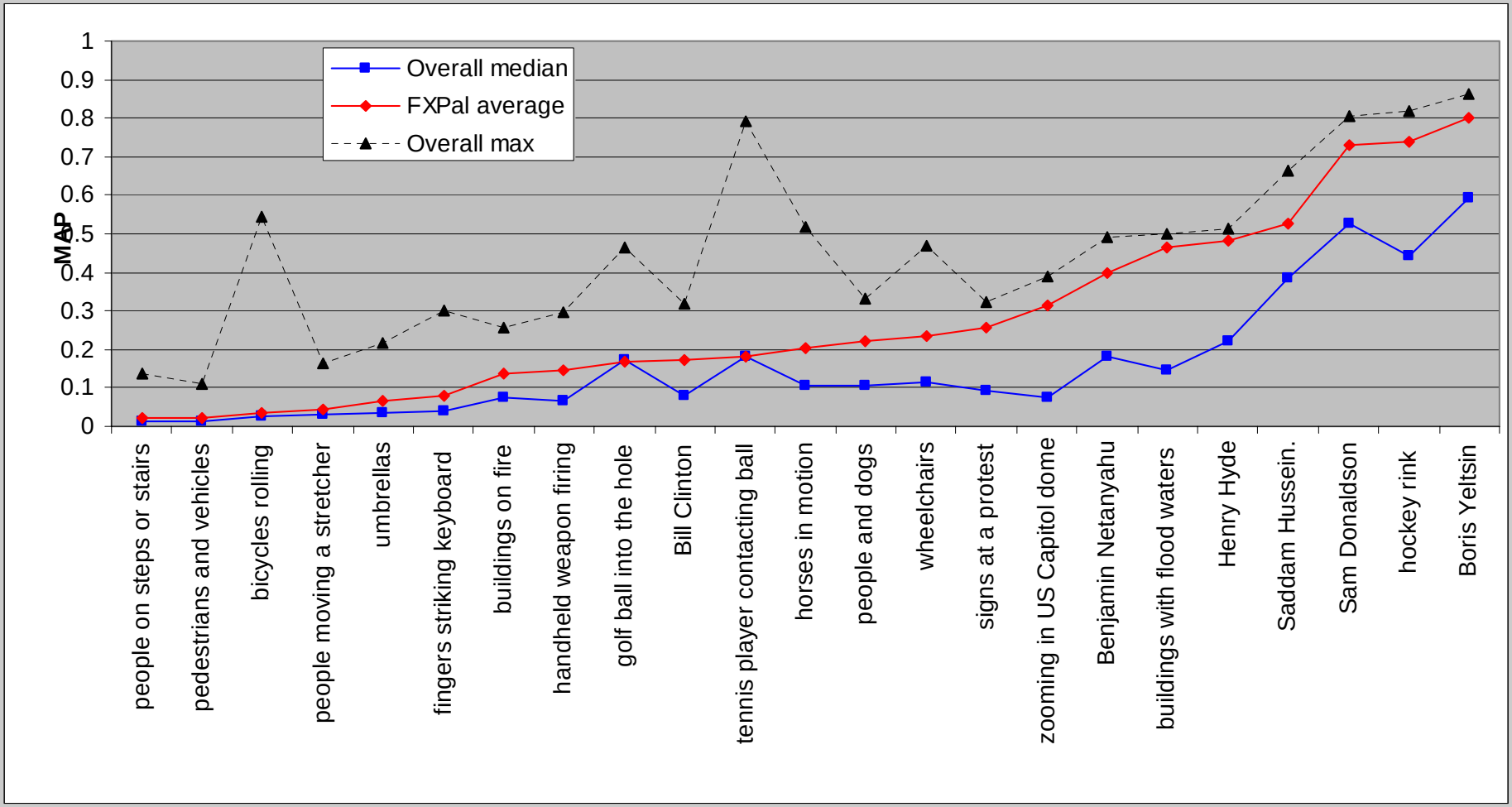
User *vs.* System



User *vs.* System in Overall



Performance by Question



Directions

- More sophisticated:
  - Story segmentation
  - Image similarity / video features
- Simplify user interface for non power-users and more typical search and re-use tasks
- Handle multiple simultaneous media streams
  - Presentation slides
  - Multi-camera capture